VISIT…

Комбинаторные формулы

Пусть имеется множество, состоящее из n элементов. Обозначим его U_n. **Перестановкой** из n элементов называется заданный порядок во множестве U_n.

Примеры перестановок:

1)распределение n различных должностей среди n человек;

2)расположение n различных предметов в одном ряду.

Сколько различных перестановок можно образовать во множестве U_n? Число перестановок обозначается P_n (читается P из n).

Чтобы вывести формулу числа перестановок, представим себе n ячеек, пронумерованных числами 1,2,...n. Все перестановки будем образовывать, располагая элементы U_n в этих ячейках. В первую ячейку можно занести любой из n элементов (иначе: первую ячейку можно заполнить n различными способами). Заполнив первую ячейку, можно n-1 способом заполнить вторую ячейку (иначе: при каждом способе заполнения первой ячейки находится n-1 способов заполнения второй ячейки). Таким образом существует $n(n$-1) способов заполнения двух первых ячеек. При заполнении первых двух ячеек можно найти n-2 способов заполнения третьей ячейки, откуда получается, что три ячейки можно заполнить $n(n$-1)(n-2) способами. Продолжая этот процесс, получим, что число способов заполнения n ячеек равно $n(n-1)(n-2)\ldots 3\cdot 2\cdot 1$. Отсюда

$$P_n = n(n - 1)(n - 2)...\cdot 3\cdot 2\cdot 1$$

Число $n(n - 1)(n - 2)...\cdot 3\cdot 2\cdot 1$, то есть произведение всех натуральных чисел от 1 до n, называется "n-факториал" и обозначается $n!$. Отсюда $P_n = n!$

Пример. $5! = 5\cdot 4\cdot 3\cdot 2 = 120$.

По определению считается: $1!=1$; $0!=1$.

Размещениями из n элементов по k элементов будем называть упорядоченные подмножества, состоящие из k элементов, множества U_n - (множества, состоящего из n элементов). Число размещений из n элементов по k элементов обозначается A_n^k (читается "A из n по k").

Примеры задач, приводящих к необходимости подсчета

1) Сколькими способами можно выбрать из 15 человек 5 кандидатов и назначить их на 5 различных должностей?

2) Сколькими способами можно из 20 книг отобрать 12 и расставить их в ряд на полке?

В задачах о размещениях полагается $k<n$. В случае, если $k=n$, то легко получить $A_n^k = P_n = n!$

Для подсчета A_n^k используем тот же метод, что использовался для подсчета P_n ,только здесь возьмем лишь k ячеек. Первую ячейку можно заполнить n способами, вторую, при заполненной первой, можно заполнить n-1 способами. Можно продолжать этот процесс до заполнения последней k-й ячейки. Эту ячейку при заполненных первых k-1 ячейках можно заполнить n-(k-1) способами (или n-k+1). Таким образом все k ячеек заполняются числом способов, равным

$$n(n - 1)(n - 2)\ldots(n - k + 2)(n - k + 1) = \frac{n!}{(n - k)!}$$

Отсюда получаем: $A_n^k = \frac{n!}{(n-k)!}$

Пример. Сколько существует различных вариантов выбора 4-х кандидатур из 9-ти специалистов для поездки в 4 различных страны?

$$A_9^4 = \frac{9!}{(9-4)!} = \frac{9!}{5!} = 9 \cdot 8 \cdot 7 \cdot 6 = 3024$$

Сочетаниями из n элементов по k элементов называются подмножества, состоящие из k элементов множества U_n (множества, состоящего из n элементов).

Одно сочетание от другого отличается только составом выбранных элементов (но не порядком их расположения, как у размещений).

Число сочетаний из n элементов по k элементов обозначается C_n^k (читается "C из n по k").

Примеры задач, приводящих к необходимости подсчета числа сочетаний:

1) Сколькими способами можно из 15 человек выбрать 6 кандидатов для назначения на работу в одинаковых должностях?

2) Сколькими способами можно из 20 книг отобрать 12 книг?

Выведем формулу для подсчета числа сочетаний. Пусть имеется множество U_n и нужно образовать упорядоченное подмножество множества U_n, содержащее k элементов (то есть образовать размещение). Делаем это так:

1) выделим какие-либо k элементов из n элементов множества U_n Это, согласно сказанному выше, можно сделать C_n^k способами;

2) упорядочим выделенные k элементов, что можно сделать $P_k = k!$ способами. Всего можно получить $C_n^k \cdot P_k$ вариантов (упорядоченных подмножеств), откуда следует: $A_n^k = C_n^k \cdot P_k$,то есть

$$C_n^k = \frac{A_n^k}{P_k} = \frac{n!}{(n-k)!\,k!}$$

Пример: 6 человек из 15 можно выбрать числом способов, равным

$$C_{15}^6 = \frac{15!}{9!\,6!} = \frac{15 \cdot 14 \cdot 13 \cdot 12 \cdot 11 \cdot 10}{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2} = 5005$$

Задачи на подсчет числа подмножеств конечного множества называются комбинаторными. Рассмотрим некоторые комбинаторные задачи.

1.Из семи заводов организация должна выбрать три для размещения трех различных заказов. Сколькими способами можно разместить заказы?

Так как все заводы различны, и из условия ясно, что каждый завод может либо получить один заказ, либо не получить ни одного, здесь нужно считать число размещений

$$A_7^3 = \frac{7!}{4!} = 7 \cdot 6 \cdot 5 = 210$$

2.Если из текста задачи 1 убрать условие различия трех заказов, сохранив все остальные условия, получим другую задачу. Теперь способ размещения заказов определяется только выбором тройки заводов, так как все эти заводы получат одинаковые заказы, и число вариантов определяется как число сочетаний.

$$C_7^3 = \frac{7!}{4! \cdot 3!} = 35$$

3.Имеются 7 заводов. Сколькими способами организация может разместить на них три различных производственных заказа? (Заказ нельзя дробить, то есть распределять его на несколько заводов).

В отличие от условия первой задачи, здесь организация может отдать все три заказа первому заводу или, например, отдать два заказа второму заводу, а один - седьмому.

Задача решается так. Первый заказ может быть размещен семью различными способами (на первом заводе, на втором и т.д.). Разместив первый заказ, имеем семь вариантов размещения второго (иначе, каждый способ размещения первого заказа может сопровождаться семью способами размещения второго). Таким образом, существует 7·7=49 способов размещения первых двух заказов. Разместив их каким-либо образом, можем найти 7 вариантов размещения третьего (иначе, каждый способ размещения первых двух заказов может сопровождаться семью различными способами распределения третьего заказа). Следовательно, существуют $49 \cdot 7 = 7^3$ способов размещения трех заказов. (Если бы заказов было n, то получилось бы 7^n способов размещения).

4.Как решать задачу 3, если в ее тексте вместо слов "различных производственных заказа" поставить "одинаковых производственных заказа"?

5.Добавим к условию задачи 1 одну фразу: организация также должна распределить три различных заказа на изготовление деревянных перекрытий среди 4-х лесопилок. Сколькими способами могут быть распределены все заказы?

Каждый из A_7^3 способов распределения заказов на заводах может сопровождаться A_4^3 способами размещения заказов на лесопилках. Общее число возможных способов размещения всех заказов будет равно

$$A_7^3 \cdot A_4^3 = \frac{7!}{4!} \cdot \frac{4!}{1!} = 7!$$

Случайный эксперимент, элементарные исходы, события.

Случайным (стохастическим) экспериментом или испытанием называется осуществление какого-либо комплекса условий, который можно практически или мысленно воспроизвести сколь угодно большое число раз.

Примеры случайного эксперимента: подбрасывание монеты, извлечение одной карты из перетасованной колоды, подсчет числа автомобилей в очереди на бензоколонке в данный момент.

Явления, происходящие при реализации этого комплекса условий, то есть в результате случайного эксперимента, называются **элементарными исходами**. Считается, что при проведении случайного эксперимента реализуется только один из возможных элементарных исходов.

Если монету подбросить один раз, то элементарными исходами можно считать выпадение герба (Г) или цифры (Ц).

Если случайным экспериментом считать троекратное подбрасывание монеты, то элементарными исходами можно считать следующие:

ГГГ, ГГЦ, ГЦГ, ЦГГ, ГЦЦ, ЦГЦ, ЦЦГ, ЦЦЦ.

Множество всех элементарных исходов случайного эксперимента называется **пространством элементарных исходов** . Будем обозначать пространство элементарных исходов буквой Ω (омега большая) *i*-й элементарный исход будем обозначатьω_i (ω-омега малая).

Если пространство элементарных исходов содержит *n* элементарных исходов, то

$$\Omega=(\omega_1, \omega_2 ,..., \omega_n).$$

Для троекратного подбрасывания монеты,

$$\Omega=(\text{ГГГ, ГГЦ,...ЦЦЦ}).$$

Если случайный эксперимент - подбрасывание игральной кости, то Ω=(1,2,3,4,5,6).

Если Ω **конечно** или **счетно**, то **случайным событием** или просто **событием** называется любое подмножество Ω**.**

Множество называется счетным, если между ним и множеством *N* натуральных чисел можно установить взаимно-однозначное соответствие.

Пример счетного множества: множество возможных значений времени прилета инопланетян на Землю, если время отсчитывать с настоящего момента и исчислять с точностью до секунды.

Примеры несчетных множеств: множество точек на заданном отрезке, множество чисел *x*, удовлетворяющих неравенству $1< x \leq 2$.

В случае несчетного множества Ω будем называть событиями только подмножества, удовлетворяющие некоторому условию (об этом будет сказано позже).

Приведем примеры событий. Пусть бросается игральная кость, и элементарным исходом считается выпавшее число очков: Ω=(1,2,3,4,5,6). *A* — событие, заключающееся в том, что выпало четное число очков: *A*=(2,4,6); *B* — событие, заключающееся в том, что выпало число очков, не меньшее 3-х: *B*=(3,4,5,6).

Говорят, что те исходы, из которых состоит событие *A*, благоприятствуют событию *A*.

События удобно изображать в виде рисунка, который называется **диаграммой Венна.** На рисунке 1 пространство элементарных исходов Ω изображено в виде прямоугольника, а множество элементарных исходов, благоприятствующих событию *A*, заключено в эллипс. Сами исходы на диаграмме Венна не изображаются, а информация о соотношении между их множествами содержится в расположении границ соответствующих областей.

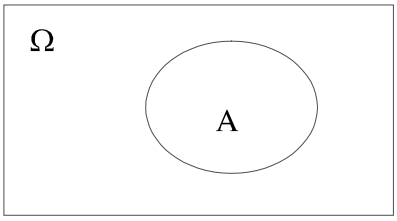

Рис.1

Суммой (объединением) двух событий *A* **и** *B* (обозначается *A*U*B*) называется событие, состоящее из всех элементарных исходов, принадлежащих по крайней мере одному из событий *A* или *B*. Событие *A*U*B* происходит, если происходит по крайней мере одно из событий *A* или *B*.

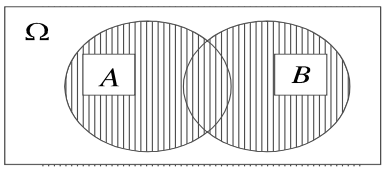

Рис.2

Приведем пример объединения событий. Пусть два стрелка стреляют в мишень одновременно, и событие *A* состоит в том, что в мишень попадает 1-й стрелок, а событие ***B*** - в том, что в мишень попадает 2-й. Событие *A*U*B* означает, что мишень поражена, или, иначе, что в мишень попал хотя бы один из стрелков.

Произведением (пересечением) $A \cap B$ событий A и B называется событие, состоящее из всех тех элементарных исходов, которые принадлежат и A и B. На рисунке 3 пересечение событий A и B изображено в виде заштрихованной области. В условиях приведенного выше примера событие $A \cap B$ заключается в том, что в мишень попали оба стрелка.

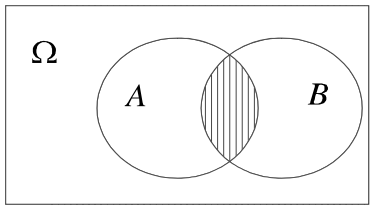

Рис.3

Разностью $A \setminus B$ или $A-B$ событий A и B называется событие, состоящее из всех исходов события A, не благоприятствующих событию B. Диаграмма Венна разности событий A и B изображена на рисунке 4.

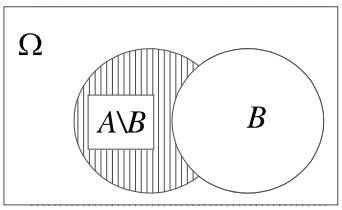

Рис.4

В условиях рассмотренного выше примера событие $A \setminus B$ заключается в том, что первый стрелок попал в мишень, а второй промахнулся.

Событие Ω называется **достоверным** (оно обязательно происходит в результате случайного эксперимента).

Пустое множество ∅ называется **невозможным** событием. Событие $\overline{A}=\Omega \setminus A$ называется **противоположным** событию A или **дополнением** события A.

События A и B называются **несовместными**, если нет исходов, принадлежащих и A и B, то есть $A \cap B = \varnothing$. На рисунке 5 изображены несовместные события A и B**.**

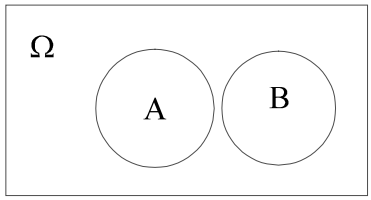

Рис.5

Непосредственно из введенных определений следуют равенства: $A \cup \overline{A}=\Omega$; $A \cap \overline{A}=\varnothing$; $\overline{A \cup B} = \overline{A} \cap \overline{B}$; $\overline{A \cap B} = \overline{A} \cup \overline{B}$. Два последних равенства называются формулами Де'Моргана.

Вероятностное пространство Случай конечного или счетного числа исходов.

Для построения полной и законченной теории случайного эксперимента или теории вероятностей, помимо введенных исходных понятий случайного эксперимента, элементарного исхода, пространства элементарных исходов, события, введем аксиому (пока для случая конечного или счетного пространства элементарных исходов).

Каждому элементарному исходу ω_i пространства Ω соответствует некоторая неотрицательная числовая характеристика P_i шансов его появления, называемая вероятностью исхода ω_i, причем

$$P_1 + P_2 + \ldots + P_n + \ldots = \sum_{i:\, \omega_i \in \Omega} P_i = 1$$

(здесь суммирование ведется по всем i, для которых выполняется условие: $\omega_i \in \Omega$).

Отсюда следует, что $0 \le P_i \le 1$ для всех i.

Вероятность любого события A определяется как сумма вероятностей всех элементарных исходов, благоприятствующих событию A. Обозначим ее $P(A)$.

$$P(A) = \sum_{i:\, \omega_i \in A} P(\omega_i) = \sum_{i:\, \omega_i \in A} P_i \qquad (*)$$

Отсюда следует, что

1) $0 \le P(A) \le 1$;

2) $P(\Omega)=1$;

3) $P(\varnothing)=0$.

Будем говорить, что задано **вероятностное пространство**, если задано пространство элементарных исходов **Ω** и определено соответствие

$$\omega_i \to P(\omega_i) = P_i.$$

Возникает вопрос: как определить из конкретных условий решаемой задачи вероятность $P(\omega_i)$ отдельных элементарных исходов?

Классическое определение вероятности.

Вычислять вероятности $P(\omega_i)$ можно, используя априорный подход, который заключается в анализе специфических условий данного эксперимента (до проведения самого эксперимента).

Возможна ситуация, когда пространство элементарных исходов состоит из конечного числа *N* элементарных исходов, причем случайный эксперимент таков, что вероятности осуществления каждого из этих *N* элементарных исходов представляются равными. Примеры таких случайных экспериментов: подбрасывание симметричной монеты, бросание правильной игральной кости, случайное извлечение игральной карты из перетасованной колоды. В силу введенной аксиомы вероятности каждого элементарного исхода в этом случае равны $\frac{1}{N}$. Из этого следует, что если событие A содержит N_A элементарных исходов, то в соответствии с определением (*)

$$P(A) = \frac{N_A}{N}$$

В данном классе ситуаций вероятность события определяется как отношение числа благоприятных исходов к общему числу всех возможных исходов.

Пример. Из набора, содержащего 10 одинаковых на вид электроламп, среди которых 4 бракованных, случайным образом выбирается 5 ламп. Какова вероятность, что среди выбранных ламп будут 2 бракованные?

Прежде всего отметим, что выбор любой пятерки ламп имеет одну и ту же вероятность. Всего существует C_{10}^5 способов составить такую пятерку, то есть случайный эксперимент в данном случае имеет C_{10}^5 равновероятных исходов.

Сколько из этих исходов удовлетворяют условию "в пятерке две бракованные лампы", то есть сколько исходов принадлежат интересующему нас событию?

Каждую интересующую нас пятерку можно составить так: выбрать две бракованные лампы, что можно сделать числом способов, равным C_4^2. Каждая пара бракованных ламп может встретиться столько раз, сколькими способами ее можно дополнить тремя не бракованными лампами, то есть $\tilde{N}_6^3$ раз. Получается, что число пятерок, содержащих две бракованные лампы, равно $C_4^2 \cdot \tilde{N}_6^3$.

Отсюда, обозначив искомую вероятность через P, получаем:

$$P = \frac{C_4^2 \cdot C_6^3}{C_{10}^5} = \frac{10}{21}$$

Статистическое определение вероятности.

Рассмотрим случайный эксперимент, заключающийся в том, что подбрасывается игральная кость, сделанная из неоднородного материала. Ее центр тяжести не находится в геометрическом центре. В этом случае мы не можем считать исходы (выпадение единицы, двойки и т.д.) равновероятными. Из физики известно, что кость более часто будет падать на ту грань, которая ближе к центру тяжести. Как определить вероятность выпадения, например, трех очков? Единственное, что можно сделать, это подбросить эту кость n раз (где n-достаточно большое число, скажем n=1000 или n=5000), подсчитать число выпадений трех очков n_3 и считать вероятность исхода, заключающегося в выпадении трех очков, равной n_3/n - относительной частоте выпадения трех очков. Аналогичным образом можно определить вероятности остальных элементарных исходов — единицы, двойки, четверки и т.д. Теоретически такой образ действий можно оправдать, если ввести **статистическое определение вероятности**.

Вероятность $P(\omega_i)$ определяется как предел относительной частоты появления исхода ω_i в процессе неограниченного увеличения числа случайных экспериментов n, то есть

$$P_i = P(\omega_i) = \lim_{n\to\infty} \frac{m_n(\omega_i)}{n},$$

где $m_n(\omega_i)$ – число случайных экспериментов (из общего числа n произведенных случайных экспериментов), в которых зарегистрировано появление элементарного исхода ω_i.

Так как здесь не приводится никаких доказательств, мы можем только надеяться, что предел в последней формуле существует, обосновывая надежду жизненным опытом и интуицией.

Геометрическая вероятность

В одном специальном случае дадим определение вероятности события для случайного эксперимента с несчетным множеством исходов.

Если между множеством Ω элементарных исходов случайного эксперимента и множеством точек некоторой плоской фигуры Σ (сигма большая) можно установить взаимно-однозначное соответствие, а также можо установить взаимно-однозначное соответствие между множеством элементарных исходов, благоприятствующих событию A, и множеством точек плоской фигуры **σ** (сигма малая), являющейся частью фигуры Σ, то

$$P(A) = \frac{s}{S},$$

где s — площадь фигуры σ, S — площадь фигуры Σ.

Пример. Два человека обедают в столовой, которая открыта с 12 до 13 часов. Каждый из них приходит в произвольный момент времени и обедает в течение 10 минут. Какова вероятность их встречи?

Пусть x — время прихода первого в столовую, а y — время прихода второго $(12 \le x \le 13;\ \ 12 \le y \le 13)$.

Можно установить взаимно-однозначное соответствие между всеми парами чисел $(x;y)$ (или множеством исходов) и множеством точек квадрата со стороной, равной 1, на координатной плоскости, где начало координат соответствует числу 12 по оси X и по оси Y, как изображено на рисунке 6. Здесь, например, точка A соответствует исходу, заключающемуся в том, что первый пришел в 12.30, а второй - в 13.00. В этом случае, очевидно, встреча не состоялась.

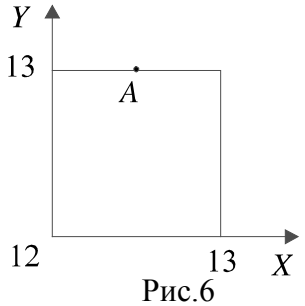

Рис.6

Если первый пришел не позже второго ($y \geq x$), то встреча произойдет при условии $0 \leq y - x \leq 1/6$ (10 мин.- это 1/6 часа).

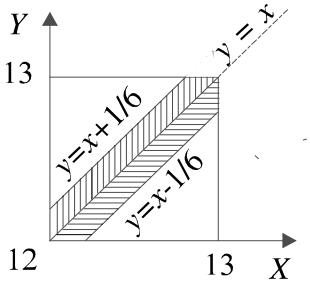

Рис. 7

Если второй пришел не позже первого ($x \geq y$), то встреча произойдет при условии $0 \leq x - y \leq 1/6$..

Между множеством исходов, благоприятствующих встрече, и множеством точек области σ, изображенной на рисунке 7 в заштрихованном виде, можно установить взаимно-однозначное соответствие.

Искомая вероятность p равна отношению площади области σ к площади всего квадрата.. Площадь квадрата равна единице, а площадь области σ можно определить как разность единицы и суммарной площади двух треугольников, изображенных на рисунке 7. Отсюда следует:

$$p = 1 - \frac{25}{36} = \frac{11}{36}$$

Непрерывное вероятностное пространство.

Как уже говорилось ранее, множество элементарных исходов может быть более, чем счетным (то есть несчетным). В этом случае нельзя считать любое подмножество множества Ω событием.

Чтобы ввести определение случайного события, рассмотрим систему (конечную или счетную) подмножеств $A_1, A_2, \ldots A_n$ пространства элементарных исходов Ω.
В случае выполнения трех условий:
1) Ω принадлежит этой системе;
2) из принадлежности A этой системе следует принадлежность $\overline{A}$ этой системе;
3) из принадлежности A_i и A_j этой системе следует принадлежность $A_i \cup A_j$ этой системе
такая система подмножеств называется **алгеброй.**

Пусть Ω — некоторое пространство элементарных исходов. Убедитесь в том, что две системы подмножеств:
1) Ω, ∅; 2) Ω, A, $\overline{A}$, ∅ (здесь A— подмножество Ω) являются алгебрами.

Пусть A_1 и A_2 принадлежат некоторой алгебре. Докажите, что $A_1 \setminus A_2$ и $A_1 \cap A_2$ принадлежат этой алгебре.

Подмножество А несчетного множества элементарных исходов Ω является событием, если оно принадлежит некоторой алгебре.

Сформулируем аксиому, называемую аксиомой А.Н. Колмогорова.

Каждому событию соответствует неотрицательное и не превосходящее единицы число *P(A)*, называемое вероятностью события *A*, причем функция *P(A)* обладает следующими свойствами:

1) $P(\Omega)=1$

2) если события $A_1, A_2,..., A_n$ несовместны, то

$$P(A_1 \cup A_2 \cup ... \cup A_n) = P(A_1) + P(A_2) + ... + P(A_n)$$

Если задано пространство элементарных исходов Ω, алгебра событий и определенная на ней функция P, удовлетворяющая условиям приведенной аксиомы, то говорят, что задано **вероятностное пространство**.

Это определение вероятностного пространства можно перенести на случай конечного пространства элементарных исходов Ω. Тогда в качестве алгебры можно взять систему всех подмножеств множества Ω.

Формулы сложения вероятностей.

Из пункта 2 приведенной аксиомы следует, что если A_1 и A_2 несовместные события, то

$$P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Если A_1 и A_2 — совместные события, то $A_1 \cup A_2 = (A_1 \setminus A_2) \cup A_2$, причем очевидно, что $A_1 \setminus A_2$ и A_2 — несовместные события. Отсюда следует:

$$P(A_1 \cup A_2) = P(A_1 \setminus A_2) + P(A_2) \qquad (*)$$

Далее очевидно: $A_1 = (A_1 \setminus A_2) \cup (A_1 \cap A_2)$, причем $A_1 \setminus A_2$ и $A_1 \cap A_2$ - несовместные события, откуда следует: $P(A_1) = P(A_1 \setminus A_2) + P(A_1 \cap A_2)$ Найдем из этой формулы выражение для $P(A_1 \setminus A_2)$ и подставим его в правую часть формулы (*). В результате получим формулу сложения вероятностей:

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

Из последней формулы легко получить формулу сложения вероятностей для несовместных событий, положив $A_1 \cap A_2 = \varnothing$.

Пример. Найти вероятность вытащить туза или червовую масть при случайном отборе одной карты из колоды в 32 листа.

$$P(\text{ ТУЗ }) = 4/32 = 1/8;\ P(\text{ ЧЕРВОВАЯ МАСТЬ }) = 8/32 = 1/4;$$

$$P(\text{ ТУЗ ЧЕРВЕЙ }) = 1/32;$$

$$P((\text{ ТУЗ }) \cup (\text{ЧЕРВОВАЯ МАСТЬ})) = 1/8 + 1/4 - 1/32 = 11/32$$

Того же результата можно было достичь с помощью классического определения вероятности, пересчитав число благоприятных исходов.

Условные вероятности.

Рассмотрим задачу. Студент перед экзаменом выучил из 30 билетов билеты с номерами с 1 по 5 и с 26 по 30. Известно, что студент на экзамене вытащил билет с номером, не превышающим 20. Какова вероятность, что студент вытащил выученный билет?

Определим пространство элементарных исходов: Ω=(1,2,3,...,28,29,30). Пусть событие A заключается в том, что студент вытащил выученный билет: A = (1,...,5,25,...,30,), а событие B — в том, что студент вытащил билет из первых двадцати: B = (1,2,3,...,20)

Событие $A \cap B$ состоит из пяти исходов: (1,2,3,4,5), и его вероятность равна 5/30. Это число можно представить как произведение дробей 5/20 и 20/30. Число 20/30 - это вероятность события B. Число 5/20 можно рассматривать как вероятность события A при условии, что событие B произошло (обозначим её $P(A/B)$). Таким образом решение задачи определяется формулой

$$P(A \cap B) = P(A/B)\, P(B)$$

Эта формула называется формулой умножения вероятностей , а вероятность $P(A/B)$ — условной вероятностью события A.

Пример..Из урны, содержащей 7 белых и 3 черных шаров, наудачу один за другим извлекают (без возвращения) два шара. Какова вероятность того, что первый шар будет белым, а второй черным?

Пусть X — событие, состоящее в извлечении первым белого шара, а Y — событие, состоящее в извлечении вторым черного шара. Тогда $X \cap Y$ - событие, заключающееся в том, что первый шар будет белым, а второй — черным. $P(Y/X)$ =3/9 =1/3 — условная вероятность извлечения вторым черного шара, если первым был извлечен белый. Учитывая, что $P(X)$ = 7/10, по формуле умножения вероятностей получаем: $P(X \cap Y)$ = 7/30

Событие А называется независимым от события В (иначе: события А и В называются независимыми), если $P(A/B)=P(A)$. За определение независимых событий можно принять следствие последней формулы и формулы умножения

$$P(A \cap B) = P(A)\, P(B)$$

Докажите самостоятельно, что если A и B — независимые события, то $\overline{A}$ и $\overline{B}$ тоже являются независимыми события.

Пример.Рассмотрим задачу, аналогичную предыдущей, но с одним дополнительным условием: вытащив первый шар, запоминаем его цвет и возвращаем шар в урну, после чего все шары перемешиваем. В данном случае результат второго извлечения никак не зависит от того, какой шар - черный или белый появился при первом извлечении. Вероятность появления первым белого шара (событие A) равна 7/10. Вероятность события B - появления вторым черного шара - равна 3/10. Теперь формула умножения вероятностей дает: $P(A \cap B)$ = 21/100.

Извлечение шаров способом, описанным в этом примере, называется **выборкой с возвращением** или **возвратной выборкой**.

Следует отметить, что если в двух последних примерах положить изначальные количества белых и черных шаров равными соответственно 7000 и 3000, то результаты расчетов тех же вероятностей будут отличаться пренебрежимо мало для возвратной и безвозвратной выборок.

Формула полной вероятности.

Пусть имеется группа событий H_1, H_2,..., H_n , обладающая следующими свойствами:

1) Все события попарно несовместны: $H_i \cap H_j = \varnothing$; $i, j = 1,2,...,n$; $i \neq j$

2) Их объединение образует пространство элементарных исходов Ω:

$\Omega = H_1 \cup H_2 \cup ... \cup H_n$.

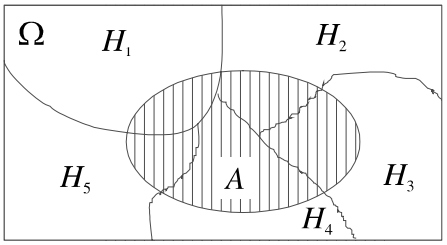

Рис.8

В этом случае будем говорить, что H_1, H_2,...,H_n образуют **полную группу событий**. Такие события иногда называют **гипотезами**.

Пусть A - некоторое событие: $A \subset \Omega$ (диаграмма Венна представлена на рисунке 8). Тогда имеет место **формула полной вероятности:**

$$P(A) = P(A/H_1)P(H_1) + P(A/H_2)P(H_2) + ... + P(A/H_n)P(H_n) = = \sum_{i=1}^{n} P(A/H_i)P(H_i)$$

Доказательство. Очевидно: $A = (A \cap H_1) \cup (A \cap H_2) \cup ... \cup (A \cap H_n)$, причем все события $A \cap H_i$ ($i = 1,2,...,n$) попарно несовместны. Отсюда по теореме сложения вероятностей получаем

$$P(A) = P(A \cap H_1) + P(A \cap H_1) + ... + P(A \cap H_n)$$

Если учесть, что по теореме умножения $P(A \cap H_i) = P(A/H_i)\,P(H_i)$ ($i = 1,2,...,n$), то из последней формулы легко получить приведенную выше формулу полной вероятности.

Пример. В магазине продаются электролампы производства трех заводов, причем доля первого завода - 30%, второго - 50%, третьего - 20%. Брак в их продукции составляет соответственно 5%, 3% и 2%. Какова вероятность того, что случайно выбранная в магазине лампа оказалась бракованной.

Пусть событие H_1 состоит в том, что выбранная лампа произведена на первом заводе, H_2 на втором, H_3 - на третьем заводе. Очевидно:

$$P(H_1) = 3/10,\ P(H_2) = 5/10,\ P(H_3) = 2/10.$$

Пусть событие A состоит в том, что выбранная лампа оказалась бракованной; A/H_i означает событие, состоящее в том, что выбранна бракованная лампа из ламп, произведенных на i-ом заводе. Из условия задачи следует:

$$P(A/H_1) = 5/10;\ P(A/H_2) = 3/10;\ P(A/H_3) = 2/10$$

По формуле полной вероятности получаем

$$P(A) = \frac{3}{10} \cdot \frac{5}{100} + \frac{5}{10} \cdot \frac{3}{100} + \frac{2}{10} \cdot \frac{2}{100} = \frac{17}{500}$$

Формула Байеса

Пусть $H_1,H_2,...,H_n$ - полная группа событий и $A \subset \Omega$ - некоторое событие. Тогда по формуле для условной вероятности

$$P(H_k / A) = \frac{P(H_k \cap A)}{P(A)} \qquad (*)$$

Здесь $P(H_k /A)$ - условная вероятность события (гипотезы) H_k или вероятность того, что H_k реализуется при условии, что событие A произошло.

По теореме умножения вероятностей числитель формулы (*) можно представить в виде

$$P(H_k \cap A) = P(A \cap H_k) = P(A /H_k)\, P(H_k)$$

Для представления знаменателя формулы (*) можно использовать формулу полной вероятности

$$P(A) = \sum_{i=1}^{n} P(A / H_i) P(H_i)$$

Теперь из (*) можно получить формулу, называемую **формулой Байеса**:

$$P(H_k / A) = \frac{P(A / H_k) P(H_k)}{\sum_{i=1}^{n} P(A / H_i) P(H_i)}$$

По формуле Байеса исчисляется вероятность реализации гипотезы H_k при условии, что событие A произошло. Формулу Байеса еще называют **формулой вероятности гипотез.**

Пример.Рассмотрим приведенную выше задачу об электролампах, только изменим вопрос задачи. Пусть покупатель купил электролампу в этом магазине, и она оказалась бракованной. Найти вероятность того, что эта лампа изготовлена на втором заводе.

Выпишем формулу Байеса для этого случая

$$P(H_2 / A) = \frac{P(A / H_2) P(H_2)}{P(A)}$$

Из этой формулы получаем: $P(H_2 / A) = 15/34$

Предлагаем читателю решить самостотельно две задачи.

.№1.В первой урне 7 белых и 3 черных шара, во второй - 8 белых и 2 черных. Из первой урны случайным образом извлекается шар и перекладывается во вторую урну. После перемешивания шаров во второй урне из нее извлекается один шар. Найти вероятность того, что извлеченный из второй урны шар — белый.

№2.В условие задачи №1 внесем изменение. Пусть после перекладывания шара из первой урны во вторую из второй урны извлечен белый шар. Найти вероятность того, что из первой урны во вторую был переложен черный шар.

Повторные независимые испытания. Формула Бернулли.

Рассмотрим случай многократного повторения одного и того же испытания или случайного эксперимента. Результат каждого испытания будем считать не зависящим от того, какой результат наступил в предыдущих испытаниях. В качестве результатов или элементарных исходов каждого отдельного испытания будем различать лишь две возможности:

1) появление некоторого события A;

2) появление события $\overline{A}$, (события, являющегося дополнением A)

Пусть вероятность $P(A)$ появления события A постоянна и равна p ($0<p<1$). Вероятность $P(\overline{A})$ события $\overline{A}$ обозначим через q: $P(\overline{A}) = 1- p=q$.

Примерами таких испытаний могут быть:

1) подбрасывание монеты: A - выпадение герба; $\overline{A}$ - выпадение цифры.

$$P(A) = P(\overline{A}) = 0{,}5.$$

2) бросание игральной кости: А - выпадение количества очков, равного пяти, $\overline{A}$ выпадение любого количества очков кроме пяти.

$$P(A) =1/6,\ P(\overline{A}) =5/6.$$

3) извлечение наудачу из урны, содержащей 7 белых и 3 черных шара, одного шара (с возвращением): A - извлечение белого шара, $\overline{A}$ - извлечение черного шара

$$P(A) = 0{,}7;\ P(\overline{A}) = 0{,}3$$

Пусть произведено n испытаний, которые мы будем рассматривать как один сложный случайный эксперимент. Составим таблицу из n клеток, расположенных в ряд, пронумеруем клетки, и результат каждого испытания будем отмечать так: если в i-м испытании событие A произошло, то в i-ю клетку ставим цифру 1, если событие A не произошло (произошло событие $\overline{A}$), в i-ю клетку ставим 0.

Если, например, проведено 5 испытаний, и событие А произошло лишь во 2 -м и 5-м испытаниях, то результат можно записать такой последовательностью нулей и единиц: 0; 1; 0; 0; 1.

Каждому возможному результату n испытаний будет соответствовать последовательность n цифр 1 или 0, чередующихся в том порядке, в котором появляются события A и $\overline{A}$ в n испытаниях, например:

$$\underbrace{1;\ 1;\ 0;\ 1;\ 0;\ 1;\ 0;\ 0;\ \ldots\ 0;\ 1;\ 1;\ 0}_{n \text{ цифр}}$$

Всего таких последовательностей можно составить 2^n (это читатель может доказать сам).

Так как испытания независимы, то вероятность P каждого такого результата определяется путем перемножения вероятностей событий A и $\overline{A}$ в соответствующих испытаниях. Так, например, для написанного выше результата найдем

$$P = p\cdot p\cdot q\cdot p\cdot q\cdot p\cdot q\cdot q\cdot\ldots\cdot q\cdot p\cdot p\cdot q$$

Если в написанной нами последовательности единица встречается x раз (это значит, что нуль встречается n-x раз), то вероятность соответствующего результата будет $p^n q^{n-x}$ независимо от того, в каком порядке чередуются эти x единиц и n-x нулей.

Все события, заключающиеся в том, что в n испытаниях событие A произошло x раз, а событие $\bar{A}$ произошло n-x раз, являются несовместными. Поэтому для вычисления вероятности объединения этих событий (или суммы этих событий), нужно сложить вероятности всех этих событий, каждая из которых равна $p^n q^{n-x}$. Всего таких событий можно насчитать столько, сколько можно образовать различных последовательностей длины n, содержащих x цифр "1" и n-x цифр "0". Таких последовательностей получается столько, сколькими способами можно разместить x цифр "1" (или n-x цифр "0") на n местах, то есть число этих последовательностей равно $C_n^x = C_n^{n-x}$

Отсюда получается **формула Бернулли**:

$$P_n(x) = C_n^x p^x q^{n-x}$$

По формуле Бернулли рассчитывается вероятность появления события A "x"раз в n повторных независимых испытаниях, где p - вероятность появления события A в одном испытании, q - вероятность появления события $\bar{A}$ в одном испытании.

Сформулированные условия проведения испытаний иногда называются "**схемой повторных независимых испытаний**" или "**схемой Бернулли**"

Число x появления события A в n повторных независимых испытаниях называется **частотой**.

Пример. Из урны, содержащей 2 белых и 6 черных шаров наудачу выбирается с возвращением 5 раз подряд один шар. Подсчитать вероятность того, что 4 раза появится белый шар.

В приведенных выше обозначениях n=8; p=1/4; q=3/4; x=5. Искомую вероятность вычисляем по формуле Бернулли:

$$P_8(5) = C_5^4 \left(\frac{1}{4}\right)^4 \left(\frac{3}{4}\right) = \frac{15}{1024}$$

По формуле Бернулли можно подсчитать вероятности всех возможных частот: x=0,1,2,3,4,5.

Заметим, что если в этой задаче считать, что белых шаров было 20000, а черных 60000, то очевидно p и q останутся неизменными. Однако в этой ситуации можно пренебречь возвращением извлеченного шара после каждой выборки (при не слишком больших значениях x) и считать вероятности всех частот: x=0,1,2,... по формуле Бернулли.

Формула Бернулли при заданных числах p и n позволяет рассчитывать вероятность любой частоты x ($0 \le x \le n$). Возникает естественный вопрос: какой частоте будет соответствовать наибольшая вероятность?

Предположим, что такая частота существует, и попытаемся ее определить из условия, что вероятность этой частоты не меньше вероятности "предыдущей" и "последующей" частот:

$$P_n(x) \ge P_n(x\text{-}1);\ P_n(x) \ge P_n(x+1) \qquad (1)$$

Первое неравенство (*) представляется в виде:

$$C_n^x p^x q^{n-x} \geq C_n^{x-1} p^{x-1} q^{n-x+1},$$

что эквивалентно $\frac{p}{x} \leq \frac{q}{n-x+1}$ или $qx \leq pn - px + p$. Отсюда следует:

$$x \leq np + p$$

Решая второе неравенство (1), получим

$$x \geq np - q$$

Таким образом, частота, имеющая наибольшую вероятность (чем вероятнейшая частота), определяется двойным неравенством

$$np - q \leq x \leq np + p$$

Если $np + p$ – целое число (тогда и $np - q$ – целое число), то две частоты: $x=np - q$ и $x=np + p$ обладают наибольшей вероятностью. Например, при $n = 7;\ p = \frac{1}{2}$, наивероятнейшие частоты: $x = 3$; $x = 4$.

Случайная величина, распределенная по закону Бернулли.

При двух заданных числах:
1) n - количестве повторных независимых испытаний,
2) p - вероятности события А в одном испытании
можно по формуле Бернулли подсчитать значение вероятности каждого целого числа x $(0 \leq x \leq n)$, где x – число появлений события А в n испытаниях (частота появления события A).

Таким образом, каждому исходу случайного эксперимента, заключающегося в серии из n испытаний по схеме Бернулли, соответствует определенное число x, рассматриваемое как случайная величина, принимающая значения 0, 1, 2,...n. Соответствие между значениями x и их вероятностями (рассчитанными по формуле Бернулли) называется законом распределения Бернулли. Строгое определение случайной величины и закона распределения будет дано позже.

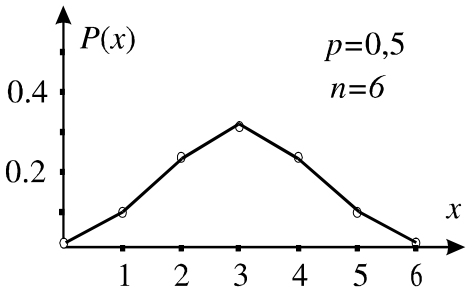

Рис. 9

Можно построить график закона распределения Бернулли (зависимости $P_n(x)$) для конкретных значений n и p. Так как аргумент x принимает лишь целые значения, график представляется в виде точек на плоскости $(x, P_n(x))$. Для наглядности точки соединяются ломаной линией, и такой график называется полигоном распределения.

При $p = 0{,}5$, как показано на рисунке 9, полигон симметричен относительно прямой $x=np$ (если p близко к 0,5, то полигон близок к симметричному)

При малых p полигон существенно асимметричен, и наивероятнейшими являются частоты, бизкие к нулю. На рисунке 10 изображен полигон распределения для p=0,2 при числе испытаний n,равном 6-ти.

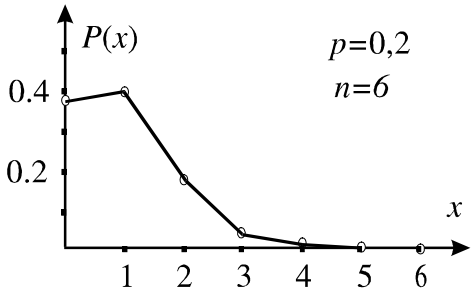

Рис.10

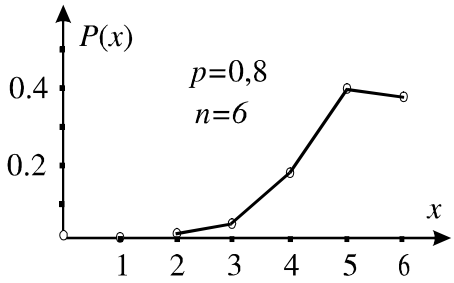

Рис. 11

При больших p, близких к 1, наиболее вероятны максимальные значения. На рисунке 11 показан полигон распределения, для p=0,8 и n=6.

О других свойствах бернуллиевского распределения будет говориться позже.

Асимптотические формулы для формулы Бернулли.

В практических задачах часто приходится вычислять вероятности различных событий, связанных с числом успехов в n испытаниях при больших значениях n. В этих случаях вычисления по формуле по формуле Бернулли становятся затруднительными. Трудности возрастают, когда приходится суммировать вероятности $P_n(x)$. К суммированию сводится вычисление вероятностей событий вида $k \le x \le l$, как, например, в такой задаче:

Проводится 70 испытаний по схеме Бернулли с вероятностью появления события А в одном испытании, равной 0,4. Найти вероятность того, что событие А произойдет от 25 до 35 раз, то есть найти $P_n(25 \le x \le 35)$.

В отдельных случаях при больших n удается заменить формулу Бернулли приближенными формулами. Такие формулы, которые получаются при условии $n \to \infty$ называются асимптотическими.

Если n достаточно велико, а p - величина очень малая, для формулы Бернулли имеет место приближенная (асимптотическая) формула

$$P_n(x) = c_n^x p^x q^{n-x} \approx \frac{\lambda^x}{x!} e^{-1}$$

Здесь $\lambda = np$ (λ - греческая буква "лямбда"). Эта формула называется **формулой Пуассона**. По формуле Пуассона вычисляются вероятности числа появлений очень редких событий в массовых испытаниях.

Задача. Телефонная станция обслуживает 1000 абонентов. В течение часа любой абонент независимо от остальных может сделать вызов с вероятностью 0,05. Требуется найти вероятность того, что в течение часа было не более 7 вызовов.

Здесь $\lambda = np = 5$. Пусть x - число вызовов. Нас интересуют значения x, равные $0, 1, \ldots, 7$.

$$P(0) = \frac{5^0}{0!} e^{-5}; P(1) = \frac{5}{1!} e^{-1}; \ldots P(7) = \frac{5^7}{7!} e^{-7}$$

$$P(0 \le x \le 7) = e^{-5}\left(1 + 5 + \frac{5^2}{2} + \frac{5^3}{6} + \frac{5^4}{24} + \frac{5^5}{120} + \frac{5^6}{720} + \frac{5^7}{5040}\right) \approx 0{,}867$$

Если n достаточно велико, p не сильно отличается от 0,5, имеет место формула Муавра-Лапласа, иногда называемая **локальной формулой Лапласа**.

$$P_n(x) = c_n^x p^x q^{n-x} = \frac{1}{\sqrt{2\pi npq}} e^{\frac{-t^2}{2}}, \qquad \text{где } t = \frac{x-np}{\sqrt{npq}}$$

Из формулы видно, что одинаковые отклонения от величины np вправо и влево здесь имеют одинаковые вероятности. В формуле Бернулли это имеет место лишь при p=0.5.

Чтобы определить вероятность того, что в 50 испытаниях по схеме Бернулли при p=0.45 событие A наступило 30 раз, нужно воспользоваться таблицей значений функции $y = e^x$. Часто встречаются таблицы значений так называемой "локальной" функции Лапласа.

$$y = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}$$

Если n достаточно велико, а p не сильно отличается от 0,5, имеет место интегральная формула Лапласа:

$$P_n(m_1 \le x \le m_2) = \sum_{x=m_1}^{m_2} c_n^x p^x q^{n-x} = \Phi(\beta) - \Phi(\alpha)$$

Здесь $\beta = \frac{x_2 - pn}{\sqrt{npq}}$; $\alpha = \frac{x_1 - np}{\sqrt{npq}}$; $\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{\frac{-u^2}{2}} du$ — функция Лапласа, значения которой определяются из таблиц.

Для вычислений используются свойства функции Лапласа

1) $\Phi(0) = 0$

2) $\Phi(\infty) = 0{,}5$

3) $\Phi(-t) = -\Phi(t)$.

При t=3,5 $\Phi(t) = 0{,}499767$, и так как $\Phi(t)$ - монотонно возрастающая функция, в практических расчетах при $t > 3{,}5$ можно принимать $\Phi(t) = 0{,}5$.

Задача. Игральную кость бросают 800 раз. Какова вероятность того, что число очков, кратное 3, выпадает не менее 280 и не более 294 раз?

Здесь $n = 800; p = \frac{1}{3}; q = \frac{2}{3}$

$$P_{300}(280 \le x \le 294) = \Phi\left(\frac{294 - 800 \cdot \frac{1}{3}}{\sqrt{800 \cdot \frac{1}{3} \cdot \frac{2}{3}}}\right) - \Phi\left(\frac{280 - 800 \cdot \frac{1}{3}}{\sqrt{800 \cdot \frac{1}{3} \cdot \frac{2}{3}}}\right) =$$

$$\Phi(2{,}05) - \Phi(1) = 0{,}479818 - 0{,}341343 \approx 0{,}14$$

Дискретные случайные величины.

Часто результатом случайного эксперимента является число. Например, можно подбросить игральную кость и получить одно из чисел: 1,2,3,4,5,6. Можно подъехать к бензоколонке и обнаружить определённое число автомашин в очереди. Можно выстрелить из пушки и измерить расстояние от места выстрела до места падения снаряда. В таких случаях будем говорить, что имеем дело со случайной величиной.

Каждому исходу случайного эксперимента поставим в соответствие единственное число x_k — значение случайной величины. Тогда **естественно рассматривать случайную величину как функцию, заданную на множестве исходов случайного эксперимента.**

Случайная величина, которая может принимать лишь конечное или счётное число значений, называется **дискретной**.

Случайные величины будем обозначать буквами греческого алфавита: ξ (кси), η (эта), … Значения случайной величины будем записывать в виде конечной или бесконечной последовательности $x_1, x_2,\ldots, x_n,\ldots$

Если говорится, что задана случайная величина ξ, это значит, что каждому исходу ω_k случайного эксперимента поставлено в соответствие единственное число x_k, что записывается в виде равенства $x_k = \xi(\omega_k)$.

Некоторые из значений x_k могут совпадать, то есть различным исходам ω может соответствовать одно и то же число x. Если все значения случайной величины совпадают, то будем говорить, что случайная величина постоянна.

Пусть A_k — множество всех элементарных исходов, каждому из которых соответствует значение x_k ($k = 1,2,\ldots,n$) случайной величины ξ. Этот факт можно записать в виде формулы

$$A_k = \bigcup_{i:\xi(\omega_i)=x_k} \omega_i$$

Таким образом, A_k – это событие (строго говоря, это верно лишь в случае конечного или счётного числа исходов). Для каждого события A_k определим число $p_k \geq 0$, равное вероятности этого события: $p_k = P(A_k)$. Очевидно, что

$$\bigcup_{k=1}^{n} A_k = \Omega,\ A_i \cap A_j = \varnothing\ \ (i,j = 1,2,\ldots,n,\ i \neq j),\ \ \sum_{i=1}^{n} p_k = 1.$$

Теперь каждому значению x_k случайной величины ξ можно поставить в соответствие вероятность $p_k = P(A_k)$ события A_k. Если такое соответствие определено то будем говорить, что задан **закон распределения** дискретной случайной величины ξ. Обычно закон распределения дискретной случайной величины представляется в виде таблицы

ξ	x_1	x_2	x_3	…	x_n
P	p_1	p_2	p_3	…	p_n

(1)

В дальнейшем для краткости будем называть величину p_i вероятностью значения x_i случайной величины. Отметим, что закон распределения содержит всю информацию о случайной величине, и задать случайную величину можно, просто представив её закон распределения.

Пусть две случайные величины

$$\xi = \{x_1,x_2,\ldots,x_n\};\ \eta = \{y_1, y_2,\ldots,y_m\} \qquad (2)$$

определены на одном и том же пространстве элементарных исходов. Если A_i (i = 1,2,…,n) – событие, объединяющее все исходы, приводящие к значению x_i случайной величины ξ, а B_j (j = 1,2,…,m) – событие, объединяющее все исходы, приводящие к значению y_i случайной величины η, то можно определить случайную величину $\zeta = \xi + \eta$, которая принимает все возможные значения $z_i^j = x_i + y_j$. Каждому такому значению z_i^j случайной величины ζ ставится в соответствие вероятность p_i^j, равная вероятности пересечения событий A_i и B_j:

$$p_i^j = P(A_i \cap B_j).$$

Таким образом определяется закон распределения суммы двух случайных величин. Также можно определить законы распределения разности $\xi - \eta$, произведения $\xi\eta$ и частного $\frac{\xi}{\eta}$ случайных величин (последний лишь в случае, если η не принимает нулевого значения).

Две случайные величины

$$\xi = \{x_1,x_2,\ldots,x_n\};\ \eta = \{y_1, y_2,\ldots,y_m\},$$

определённые на одном и том же пространстве элементарных исходов, имеющие законы распределения

ξ	x_1	…	x_i	…
P	p_1^1	…	p_i^1	…

η	y_1	…	y_j	…
P	p_1^2	…	p_j^2	…

называются **независимыми**, если при любых i и j выполняется равенство

$$P((\xi = x_i) \cap (\eta = y_j)) = p_i^1\ p_j^2$$

Пример1. Брошены две игральных кости. Число очков, выпавшее на первой кости, – случайная величина ξ. Число очков, выпавшее на второй кости

– случайная величина η. Считаем, что все исходы $((\xi = i) \cap (\eta = j))$ ($i = 1,2,\ldots,6$; $j = 1,2, \ldots,6$) равновероятны, всего их 36, поэтому

$$P((\xi = i) \cap (\eta = j)) = \frac{1}{36}$$

Так как $P(\xi = i) = \frac{1}{6}$ и $P(\eta = j)) = \frac{1}{6}$, очевидно, что по определению ξ и η – независимые случайные величины.

Пример 2. Даны две независимые случайные величины ξ и η с заданными законами распределения

ξ	0	1
P	$\frac{1}{3}$	$\frac{2}{3}$

η	1	2
P	$\frac{1}{4}$	$\frac{3}{4}$

Определим случайные величины α и β следующим образом: $\alpha = \xi + \eta$, $\beta = \xi\eta$. Выясним, являются ли независимыми случайные величины α и β.

Составим закон распределения α. Наименьшее значение α равняется 1. Вероятность события $\alpha = 1$ равна вероятности события $(\xi = 0) \cap (\eta = 1)$, которая в силу независимости ξ и η равна $\frac{1}{3} \cdot \frac{1}{4} = \frac{1}{12}$. Событие $\alpha = 2$ совпадает с событием $((\xi = 0) \cap (\eta = 2)) \cup ((\xi = 1) \cap (\eta = 1))$. Его вероятность равна

$$\frac{1}{3} \cdot \frac{3}{4} + \frac{2}{3} \cdot \frac{1}{4} = \frac{5}{12}.$$

Максимальное значение α, равное 3, имеет вероятность $\frac{1}{2}$. Таким образом, закон распределения случайной величины α можно представить таблицей

α	1	2	3
P	$\frac{1}{12}$	$\frac{5}{12}$	$\frac{6}{12}$

Закон распределения β представляется таблицей

β	0	1	2
P	$\frac{1}{3}$	$\frac{5}{12}$	$\frac{1}{2}$

Рассмотрим события $\alpha = 3$ и $\beta = 0$. Очевидно, что

$$P(\alpha = 3)\, P(\beta = 0) = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}$$

С другой стороны, событие $(\alpha = 3) \cap (\beta = 0)$ – невозможное, так как $\alpha = 3$ только при $\xi = 1$, а $\beta = 0$ лишь при $\xi = 0$. Отсюда следует, что

$$P((\alpha = 3) \cap (\beta = 0)) = 0,$$

и теперь ясно, что, по крайней мере, в одном случае условие определения независимости для случайных величин α и β не выполняется. Отсюда следует, что эти случайные величины зависимы.

Математическое ожидание случайной величины.

Пусть задан закон распределения случайной величины ξ.

ξ	x_1	x_2	x_3	…	x_n
P	p_1	p_2	p_3	…	p_n

Математическое ожидание $M\xi$ (или $M(\xi)$) случайной величины ξ определяется формулой

$$M\xi = \sum_{i=1}^{n} x_i p_i$$

Рассмотрим пример. Пусть в некотором магазине, торгующем электробытовой техникой, получены статистические данные о числе проданных холодильников в каждый день месяца (условно считаем, что месяц состоит из 30 рабочих дней). Эти данные собраны в таблицу

Количество проданных холодильников	0	1	2	3	4	5
Число дней, в которые было продано столько холодильников	3	7	8	9	2	1

По этой таблице легко подсчитать число холодильников, проданных в магазине за месяц: 0*1+1*7+2*8+3*9+4*2+5*1 = 63. Чтобы подсчитать среднее число холодильников, продававшихся в один день месяца, нужно эту сумму разделить на 30, в результате получим 2,1. Если в приведенной таблице каждое число второй строки поделить на 30, то получится последовательность дробей

$$\frac{1}{10}; \frac{7}{30}; \frac{4}{15}; \frac{3}{10}; \frac{1}{15}; \frac{1}{30},$$

каждая из которых представляет собой так называемую **относительную частоту**, с которой в данный месяц появлялся приведенный в верхней строке объём продаж. Очевидно, что если просуммировать все произведения чисел, стоящих в первой строке таблицы, на их относительные частоты, то получится то же среднее число продававшихся в один день холодильников:

$$0 \cdot \frac{1}{10} + 1 \cdot \frac{7}{30} + 2 \cdot \frac{4}{15} + 3 \cdot \frac{3}{10} + 4 \cdot \frac{1}{15} + 5 \cdot \frac{1}{30} = 2{,}1$$

Если бы в последней формуле относительные частоты рассчитывались не для одного месяца, а для существенно большего срока, то при некоторых условиях (например, при отсутствии кризисных явлений, существенно влияющих на спрос населения на дорогостоящие товары) эти относительные частоты можно было бы считать довольно близкими к вероятностям соответствующих значений объёма продаж. Таким образом, приходим к выводу, что математическое ожидание случайной величины – это в некотором смысле её среднее значение. Следует отметить, что случайная величина может вообще не принимать значения, равного её математическому ожиданию. Так, например, случайная величина, принимающая только значения 1 и –1, каждое – с вероятностью 0,5, имеет математическое ожидание, равное нулю.

Пример. Найти математическое ожидание случайной величины, заданной законом распределения

ξ	1	0
P	p	q

Здесь p + q = 1.

$$M\xi = 1 \cdot p + 0 \cdot q = p$$

Свойства математического ожидания.

1. Если случайная величина ξ принимает одно и то же значение при всех исходах случайного эксперимента, то есть $\xi \equiv C$, то её математическое ожидание равно C.
2. Если $M\xi = a$, и k – константа, то $M(k\xi) = kM\xi$ (математическое ожидание случайной величины, умноженной на число, равно математическому ожиданию случайной величины, умноженному на это число).
3. Если $M\xi = a$, и k – константа, то $M(k + \xi) = k + M\xi$ (математическое ожидание суммы случайной величины и числа равно сумме этого числа и математического ожидания случайной величины).

Выведем формулу для математического ожидания суммы двух случайных величин ξ и η, определённых на одном и том же пространстве элементарных исходов и заданных законами распределения

ξ	x_1	…	x_n
P	p_1^1	…	p_n^1

η	y_1	…	y_k
P	p_1^2	…	p_k^2

$M(\xi + \eta) = (x_1 + y_1)P((\xi = x_1) \cap (\eta = y_1)) + (x_2 + y_1)P((\xi = x_2) \cap (\eta = y_1)) + \ldots$
$+(x_i + y_j)P((\xi = x_i) \cap (\eta = y_j)) + \ldots + (x_n + y_k)P((\xi = x_n) \cap (\eta = y_k))$

Очевидно, что сумма в правой части последней формулы содержит nk слагаемых. Преобразуем эту сумму следующим образом:

$$M(\xi+\eta) = x_1 P((\xi=x_1)\cap(\eta=y_1)) + x_1 P((\xi=x_1)\cap(\eta=y_2)) +\ldots+x_1 P((\xi=x_1)\cap(\eta=y_k)) +$$
$$+ x_2P((\xi=x_2)\cap(\eta=y_1)) + x_2P((\xi=x_2)\cap(\eta=y_2)) +\ldots + x_2P((\xi=x_2)\cap(\eta=y_k)) + \ldots$$
$$+ x_nP((\xi=x_n)\cap(\eta=y_1)) + x_nP((\xi=x_n)\cap(\eta=y_2)) +\ldots + x_nP((\xi=x_n)\cap(\eta=y_k)) +$$
$$+ y_1P((\xi=x_1)\cap(\eta=y_1)) + y_1P((\xi=x_2)\cap(\eta=y_1)) +\ldots + y_1P((\xi=x_n)\cap(\eta=y_1)) +$$
$$+ y_2P((\xi=x_1)\cap(\eta=y_2)) + y_2P((\xi=x_2)\cap(\eta=y_2)) +\ldots + y_2P((\xi=x_n)\cap(\eta=y_2)) + \ldots$$
$$+ y_kP((\xi=x_1)\cap(\eta=y_k)) + y_kP((\xi=x_2)\cap(\eta=y_k)) +\ldots + y_kP((\xi=x_n)\cap(\eta=y_k)) =$$
$$= x_1(P((\xi=x_1)\cap(\eta=y_1)) + P((\xi=x_1)\cap(\eta=y_2)) +\ldots + P((\xi=x_1)\cap(\eta=y_k))) +$$
$$+ x_2(P((\xi=x_2)\cap(\eta=y_1)) + P((\xi=x_2)\cap(\eta=y_2)) +\ldots + P((\xi=x_2)\cap(\eta=y_k))) +\ldots +$$
$$+ x_n(P((\xi=x_n)\cap(\eta=y_1)) + P((\xi=x_n)\cap(\eta=y_2)) +\ldots + P((\xi=x_n)\cap(\eta=y_k))) +$$
$$+ y_1(P((\xi=x_1)\cap(\eta=y_1)) + P((\xi=x_2)\cap(\eta=y_1)) +\ldots + P((\xi=x_n)\cap(\eta=y_1))) +$$
$$+ y_2(P((\xi=x_1)\cap(\eta=y_2)) + P((\xi=x_2)\cap(\eta=y_2)) +\ldots + P((\xi=x_n)\cap(\eta=y_2))) + \ldots$$
$$+ y_k(P((\xi=x_1)\cap(\eta=y_k)) + P((\xi=x_2)\cap(\eta=y_k)) +\ldots + P((\xi=x_n)\cap(\eta=y_k))) =$$
$$= x_1P(\xi=x_1) + x_2P(\xi=x_2) +\ldots+ x_n P(\xi=x_n) +$$
$$+ y_1P(\eta=y_1) + y_2P(\eta=y_2) +\ldots+ y_1P(\eta=y_1) = M\xi + M\eta$$

При выводе этой формулы использован очевидный факт, что, например, событие $\xi=x_1$ можно представить в виде объединения несовместных событий $(\xi=x_1)\cap(\eta=y_1)$, $(\xi=x_1)\cap(\eta=y_2)$, …, $(\xi=x_1)\cap(\eta=y_n)$.

Пример.

Заданы n одинаково распределённых случайных величин ξ_1, ξ_2, …, ξ_n с законом распределения

ξ_i	1	0
P	p	q

Найти математическое ожидание суммы этих случайных величин.

Решение.

$$M(\sum_{i=1}^{n}\xi_i) = \sum_{i=1}^{n} M\xi_i = np$$

Теорема.

Если случайные величины ξ и η независимы, то

$$M(\xi\eta) = M\xi\cdot M\eta$$

Доказательство.

Если заданы законы распределения двух независимых случайных величин ξ и η

ξ	x_1	…	x_i	…	x_n
P	p_1^1	…	p_i^1	…	p_n^1

η	y_1	…	y_j	…	y_k
P	p_1^2	…	p_j^2	…	p_k^2

то математическое ожидание произведения этих случайных величин можно представить следующим образом:

$$M(\xi\eta) = \sum_{i=1}^{n}\sum_{j=1}^{k} x_i y_j p_i^1 p_j^2 =$$

$$= x_1 p_1^1 \sum_{j=1}^{k} y_j p_j^2 + x_2 p_2^1 \sum_{j=1}^{k} y_j p_j^2 + \ldots + x_i p_i^1 \sum_{j=1}^{k} y_j p_j^2 \ldots + x_n p_n^1 \sum_{j=1}^{k} y_j p_j^2 =$$

$$= x_1 p_1^1 M\eta + x_2 p_2^1 M\eta + \ldots + x_i p_i^1 M\eta \ldots + x_n p_n^1 M\eta = M\eta \sum_{i=1}^{n} x_i p_i = M\xi \cdot M\eta$$

Дисперсия случайной величины.

Дисперсия $D\xi$ случайной величины ξ определяется формулой

$$D\xi = M(\xi - M\xi)^2$$

Дисперсия случайной величины — это математическое ожидание квадрата отклонения случайной величины от её математического ожидания.

Рассмотрим случайную величину ξ с законом распределения

ξ	1	2	3
P	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

Вычислим её математическое ожидание.

$$M\xi = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{2} + 3\cdot\frac{1}{3} = \frac{13}{6}$$

Составим закон распределения случайной величины $\xi - M\xi$

$\xi - M\xi$	$-\frac{7}{6}$	$-\frac{1}{6}$	$\frac{5}{6}$
P	$\frac{1}{6}$	$\frac{1}{2}$	$\frac{1}{3}$

а затем закон распределения случайной величины $(\xi - M\xi)^2$

$(\xi - M\xi)^2$	$\frac{1}{36}$	$\frac{25}{36}$	$\frac{49}{36}$
P	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Теперь можно рассчитать величину $D\xi$:

$$D\xi = \frac{1}{36}\cdot\frac{1}{2} + \frac{25}{36}\cdot\frac{1}{3} + \frac{49}{36}\cdot\frac{1}{6} = \frac{17}{36}$$

Используя определение дисперсии, для дискретной случайной величины формулу вычисления дисперсии можно представить в таком виде:

$$D\xi = \sum_{i=1}^{n}\left(x_i - M\xi\right)^2 p_i$$

Можно вывести ещё одну формулу для вычисления дисперсии:

$$D\xi = \sum_{i=1}^{n}\left(x_i - M\xi\right)^2 p_i = \sum_{i=1}^{n}\left(x_i{}^2 - 2x_i M\xi + M^2\xi\right)p_i =$$

$$= \sum_{i=1}^{n} x_i{}^2 p_i - 2M\xi\sum_{i=1}^{n} x_i p_i + M^2\xi\sum_{i=1}^{n} p_i = M\xi^2 - 2M\xi\cdot M\xi + M^2\xi =$$

$$= M\xi^2 - M^2\xi$$

Таким образом, дисперсия случайной величины равна разности математического ожидания квадрата случайной величины и квадрата её математического ожидания.

Пример.

Найти дисперсию случайной величины, заданной законом распределения

ξ	1	0
P	p	q

Выше было показано, что $M\xi = p$. Легко видеть, что $M\xi^2 = p$. Таким образом, получается, что $D\xi = p - p^2 = pq$.

Дисперсия характеризует степень рассеяния значений случайной величины относительно её математического ожидания. Если все значения случайной величины тесно сконцентрированы около её математического ожидания и большие отклонения от математического ожидания маловероятны, то такая случайная величина имеет малую дисперсию. Если значения случайной величины рассеяны и велика вероятность больших отклонений от математического ожидания, то такая случайная величина имеет большую дисперсию.

Свойства дисперсии.

1. Если k – число, то $D(k\xi) = k^2 D\xi$.

Доказательство.

$$D(k\xi) = M(k\xi - M(k\xi))^2 = M(k\xi - k\,M\xi)^2 = M(k^2\,(\xi - M\xi)^2) = k^2 M(\xi - M\xi)^2 =$$

$$= k^2\,D\xi$$

2. Для попарно независимых случайных величин $\xi_1, \xi_2, \ldots, \xi_n$ справедливо равенство

$$D\sum_{i=1}^{n}\xi_i = \sum_{i=1}^{n} D\xi_i$$

Это свойство оставим без доказательства. Рекомендуем читателю рассмотреть следующий пример.

Пусть ξ и η – независимые случайные величины с заданными законами распределения:

ξ	0	1
P	0,25	0,75

η	1	2
P	0,7	0,7

Показать, что $D(\xi + \eta) = D\xi + D\eta$.

Биномиальный закон распределения.

Пусть заданы числа $n \in N$ и p $(0 \le p \le 1)$. Тогда каждому целому числу из промежутка $[0; n]$ можно поставить в соответствие вероятность, рассчитанную по формуле Бернулли. Получим закон распределения случайной величины (назовём её β)

β	0	…	k	…	n
P	…	…	$C_n^k p^k (1-p)^{n-k}$	…	…

Будем говорить, что случайная величина β распределена по закону Бернулли. Такой случайной величиной является частота появления события A в n повторных независимых испытаниях, если в каждом испытании событие A происходит с вероятностью p.

Рассмотрим отдельное i-е испытание. Пространство элементарных исходов для него имеет вид

$$\Omega = \left\{A, \overline{A}\right\}$$

Определим на этом пространстве случайную величину ξ_i следующим образом:

$\xi_i = 1$, если происходит событие A;

$\xi_i = 0$, если происходит событие $\overline{A}$

Закон распределения случайной величины ξ_i рассматривался в предыдущем параграфе.

ξ_i	1	0
P	p	$q = 1–p$

$$M\xi = \cdot p;\ D\xi = pq$$

Для $i = 1,2,\ldots,n$ получаем систему из n независимых случайных величин ξ_i, имеющих одинаковые законы распределения. Если теперь сравнить законы распределения двух случайных величин β и $\sum_{i=1}^{n} \xi i$, то можно сделать очевидный вывод: $\beta = \sum_{i=1}^{n} \xi_i$. Отсюда следует, что для случайной величины β, имеющей

закон распределения Бернулли, математическое ожидание и дисперсия определяются формулами

$$M\beta = M\sum_{i=1}^{n}\xi_i = \sum_{i=1}^{n}M\xi_i = \sum_{i=1}^{n}p = np;$$

$$D\beta = D\sum_{i=1}^{n}\xi_i = \sum_{i=1}^{n}D\xi_i = \sum_{i=1}^{n}pq = npq$$

Найдём оценку величины p — вероятности успеха в одном испытании некоторого биномиального эксперимента. Для этого проведём n испытаний и подсчитаем x – число успехов. Оценку p^* неизвестной величины p определим формулой $p^* = \frac{x}{n}$.

Пример.

Из 20 отобранных для контроля образцов продукции 4 оказались нестандартными. Оценим вероятность того, что случайно выбранный экземпляр продукции не отвечает стандарту отношением $p^* = 4/20 = 0{,}2$.

Так как x случайная величина, p^* – тоже случайная величина. Значения p^* могут меняться от одного эксперимента к другому (в рассматриваемом случае экспериментом является случайный отбор и контроль 20-ти экземпляров продукции). Каково математическое ожидание p^*? Поскольку x есть случайная величина, обозначающая число успехов в n испытаниях по схеме Бернулли, $Mx = np$. Для математического ожидания случайной величины p^* по определению получаем: $Mp^* = M\left(\frac{x}{n}\right)$, но n здесь является константой, поэтому по свойству математического ожидания

$$Mp^* = \frac{1}{n}Mx = \frac{1}{n}np = p$$

Таким образом, “в среднем” получается истинное значение p, чего и следовало ожидать. Это свойство оценки p^* величины p имеет название: p^* является **несмещённой** оценкой для p. Отсутствие систематического отклонения от величины оцениваемого параметра p подтверждает целесообразность использования величины p^* в качестве оценки. Вопрос о точности оценки пока оставляем открытым.

Непрерывные случайные величины.

Случайная величина, значения которой заполняют некоторый промежуток, называется непрерывной.

В частных случаях это может быть не один промежуток, а объединение нескольких промежутков. Промежутки могут быть конечными, полубесконечными или бесконечными, например: $(a; b]$, $(-\infty; a)$, $[b;\infty)$, $(-\infty; \infty)$.

Вообще непрерывная случайная величина – это абстракция. Снаряд, выпущенный из пушки, может пролететь любое расстояние, скажем, от 5 до 5,3 километров, но никому не придёт в голову измерять эту величину с точностью до 0,0000001 километра (то есть до миллиметра), не говоря уже об абсолютной точности. В практике такое расстояние будет дискретной случайной величиной, у которой одно значение от другого отличается по крайней мере на 1 метр.

При описании непрерывной случайной величины принципиально невозможно выписать и занумеровать все её значения, принадлежащие даже достаточно узкому интервалу. Эти значения образуют несчётное множество, называемое «континуум».

Если ξ – непрерывная случайная величина, то равенство $\xi = x$ представляет собой, как и в случае дискретной случайной величины, некоторое случайное событие, но для непрерывной случайной величины это событие можно связать лишь с вероятностью, равной нулю, что однако не влечёт за собой невозможности события. Так например, можно говорить, что только с вероятностью «нуль» снаряд пролетит 5245,7183 метра, или что отклонение действительного размера детали от номинального составит 0,001059 миллиметра. В этих случаях практически невозможно установить, произошло событие или нет, так как измерения величин проводятся с ограниченной точностью, и в качестве результата измерения можно фактически указать лишь границы более или менее узкого интервала, внутри которого находится измеренное значение.

Значениям непрерывной случайной величины присуща некоторая неопределенность. Например, нет практического смысла различать два отклонения от номинального размера, равные 0,5 мм и 0,5000025 мм. Вероятность, отличная от нуля, может быть связана только с попаданием

величины в заданный, хотя бы и весьма узкий, интервал. Здесь можно привести сравнение с распределением массы вдоль стержня. Отсутствует масса, сосредоточенная, скажем, в сечении, расположенном на расстоянии 20 см от левого конца стержня, имеет смысл говорить лишь о массе, заключённой между сечениями, проходящими через концы некоторого промежутка.

Пусть ξ – непрерывная случайная величина. Рассмотрим для некоторого числа x вероятность неравенства $x < \xi < x + \Delta x$

$$P(x < \xi < x + \Delta x).$$

Здесь Δx – величина малого интервала.

Очевидно, что если $\Delta x \to 0$, то $P(x < \xi < x + \Delta x) \to 0$. Обозначим $p(x)$ предел отношения $P(x < \xi < x + \Delta x)$ к при $\Delta x \to 0$, если такой предел существует:

$$\lim_{\Delta x \to 0} \frac{P(x < \xi < x + \Delta x)}{\Delta x} = p(x) \qquad (1)$$

Функция $p(x)$ называется плотностью распределения случайной величины. Из формулы (1) следует равенство, справедливое для малых величин Δx, которое также можно считать определением функции $p(x)$:

$$P(x < \xi < x + \Delta x) \cong p(x)\Delta x \qquad (2)$$

Очевидно, что $p(x)$ – неотрицательная функция. Для определения вероятности того, что случайная величина ξ примет значение из промежутка $[a, b]$ конечной длины, нужно выбрать на промежутке произвольные числа $x_1, x_2, \ldots, x_n$ удовлетворяющие условию $a=x_0<x_1<x_2<\ldots<x_n<b=x_{n+1}$. Эти числа разобьют промежуток $[a, b]$ на $n+1$ частей, представляющих собой промежутки $[x_0, x_1)$, $[x_1, x_2)$, …,$[x_n, b]$. Введём обозначения:

$$\Delta x_0 = x_1 - x_0,\ \Delta x_1 = x_2 - x_1,\ \ldots,\ \Delta x_n = b - x_n,$$

и составим сумму $\sum_{i=1}^{n} p(x_i)\Delta x_i$. Рассмотрим процесс, при котором число точек разбиения неограниченно возрастает таким образом, что максимальная величина Δx_i стремится к нулю. Будем считать функцию $p(x)$ непрерывной на

промежутке $(a; b)$, тогда пределом суммы $\sum_{i=1}^{n} p(x_i)\Delta x_i$ будет определённый интеграл по промежутку $[a; b]$ от функции $p(x)$, равный искомой вероятности:

$$P(a \le \xi \le b) = \int_a^b p(x)dx \qquad (3)$$

Это равенство можно также рассматривать как определение функции $p(x)$. Отсюда следует, что вероятность попадания случайной величины в любой интервал (x_1, x_2) равна площади фигуры, образованной отрезком $[x_1, x_2]$ оси x, графиком функции $p(x)$ и вертикальными прямыми $x = x_1$, $x = x_2$, как изображено на рисунке 1.

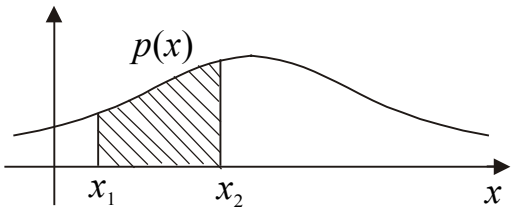

Рис. 1

Если все возможные значения случайной величины принадлежат интервалу $(a; b)$, то для $p(x)$ – её плотности распределения справедливо равенство

$$\int_a^b p(x)dx = 1$$

Для удобства иногда считают функцию $p(x)$ определённой для всех значений x, полагая её равной нулю в тех точках x, которые не являются возможными значениями этой случайной величины.

Плотностью распределения может служить любая интегрируемая функция $p(x)$, удовлетворяющая двум условиям:

1) $p(x) \ge 0$;

2) $\int_{-\infty}^{\infty} p(x)dx = 1$

Можно задавать случайную величину, задавая функцию $p(x)$, удовлетворяющую этим условиям.

В качестве примера рассмотрим случайную величину ξ, равномерно распределённую на промежутке $[a; b]$. В этом случае $p(x)$ постоянна внутри этого промежутка:

$$p(x)=\begin{cases} c & a \le x \le b \\ 0 & x < a; x > b \end{cases}$$

По свойству 2) функции $p(x)$

$$\int_{-\infty}^{\infty} p(x)dx = \int_{a}^{b} cdx = c(b-a) = 1$$

Отсюда $c = \dfrac{1}{b-a}$. График функции $p(x)$ представлен на рисунке 2.

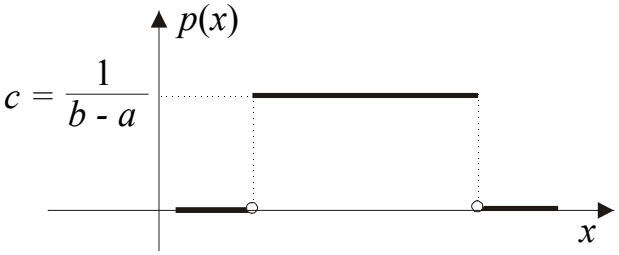

Рис. 2

Во многих практических задачах встречаются случайные величины, у которых возможные значения не ограничены сверху и снизу. В этом случае кривая распределения располагается над осью x и при $x \to \infty$ и $x \to -\infty$ асимптотически приближается к этой оси, как изображено на рисунке 1. Вероятность того, что случайная величина ξ примет значение, меньшее некоторого числа a, равна площади фигуры, заключённой между кривой распределения и горизонтальной координатной осью слева от точки a. Будем считать, что такая площадь существует.

Пусть ξ – непрерывная случайная величина. Функция $F(x)$, которая определяется равенством

$$F(x) = P(\xi \le x),$$

называется **интегральной функцией распределения** или просто **функцией распределения** случайной величины ξ. Непосредственно из определения следует равенство $F(x) = \int_{-\infty}^{x} p(t)dt$. Формула производной определённого интеграла по верхнему пределу в данном случае приводит к соотношению $F'(x) = p(x)$. Плотность распределения $p(x)$ называют **дифференциальной функцией распределения.**

Функция распределения $F(x)$ случайной величины ξ имеет следующие свойства.

1. $F(x)$ — непрерывная возрастающая функция.

2. $\lim\limits_{x\to-\infty} F(x)=0$; $\lim\limits_{x\to\infty} F(x)=1$

Свойства 1 и 2 вытекают непосредственно из определения функции $F(x)$.

3. Приращение $F(x)$ на промежутке $(x_1;\ x_2)$ равно вероятности того, что случайная величина ξ принимает значение из этого промежутка:

$$F(x_2) - F(x_1) = P(x_1 < \xi \le x_2)$$

Доказательство.

$$F(x_2) = P(\xi \le x_2) = P(\xi \le x_1) + P(x_1 < \xi \le x_2) = F(x_1) + P(x_1 < \xi \le x_2)$$

Отсюда

$$P(x_1 < \xi \le x_2) = F(x_2) - F(x_1)$$

Заметим, что для непрерывной случайной величины ξ справедливы равенства

$$P(x_1 < \xi \le x_2) = P(x_1 < \xi < x_2) = P(x_1 \le \xi < x_2) = P(x_1 \le \xi \le x_2)$$

Для равномерного распределения функция $F(x)$ имеет вид:

$$F(x)=\begin{cases} 0 & \text{при} \quad x \le a \\ \int\limits_a^x \dfrac{dt}{b-a} = \dfrac{x-a}{b-a} & \text{при } a < x < b \\ 1 & \text{при} \quad x \ge b \end{cases}$$

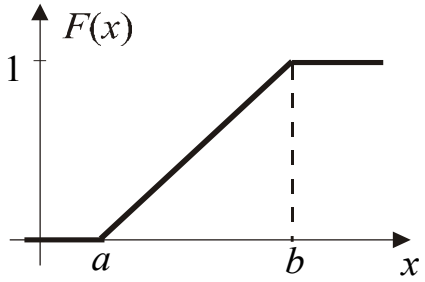

Рис. 3

График функции $F(x)$ представлен на рисунке 3.

Закон распределения непрерывной случайной величины можно определить заданием либо функции $p(x)$, либо функции $F(x)$.

Правило 3-х σ (трех “сигм”).

Пусть имеется нормально распределённая случайная величина ξ с математическим ожиданием, равным a и дисперсией σ^2. Определим вероятность попадания ξ в интервал $(a - 3\sigma; a + 3\sigma)$, то есть вероятность того, что ξ принимает значения, отличающиеся от математического ожидания не более, чем на три среднеквадратических отклонения.

$$P(a - 3\sigma < \xi < a + 3\sigma) = \Phi(3) - \Phi(-3) = 2\Phi(3)$$

По таблице находим $\Phi(3)=0{,}49865$, откуда следует, что $2\Phi(3)$ практически равняется единице. Таким образом, можно сделать важный вывод: нормальная случайная величина принимает значения, отклоняющиеся от ее математического ожидания не более чем на 3σ.

(Выбор числа 3 здесь условен и никак не обосновывается: можно было выбрать 2,8, 2,9 или 3,2 и получить тот же вероятностный результат. Учитывая, что $\Phi(2)=0{,}477$, можно было бы говорить и о правиле 2–х “сигм”.)

Совместное распределение двух случайных величин.

Пусть пространство элементарных исходов Ω случайного эксперимента таково, что каждому исходу ω_i^j ставиться в соответствие значение случайной величины ξ, равное x_i и значение случайной величины η, равное y^j.

Примеры:

1. Представим себе большую совокупность деталей, имеющих вид стержня. Случайный эксперимент заключается в случайном выборе одного стержня. Этот стержень имеет длину, которую будем обозначать ξ и толщину—η (можно указать другие параметры—объем, вес, чистота обработки, выраженная в стандартных единицах).
2. Если результат эксперимента—случайный выбор какого–либо предприятия в данной области, то за ξ можно принимать объем производства отнесенный к количеству сотрудников, а за η—объем продукции, идущей на экспорт, тоже отнесенной к числу сотрудников.

В этом случае мы можем говорить о совместном распределении случайных величин ξ и η или о “двумерной” случайной величине.

Если ξ и η дискретны и принимают конечное число значений (ξ – n значений, а η – k значений), то закон совместного распределения случайных величин ξ и η можно задать, если каждой паре чисел x_i, y^j (где x_i принадлежит множеству значений ξ, а y^j—множеству значений η) поставить в соответствие вероятность p_i^j, равную вероятности события, объединяющего все исходы ω_i^j (и состоящего лишь из этих исходов), которые приводят к значениям

$$\xi = x_i;\ \eta = y^j.$$

Такой закон распределения можно задать в виде таблицы:

η / ξ	y^1	y^2	…	y^j	…	y^k	
x_1	$p_1^{\,1}$	$p_1^{\,2}$	…	$p_1^{\,j}$	…	$p_1^{\,k}$	P_1
…	…	…	…	…	…	…	…
x_i	$p_i^{\,1}$	$p_i^{\,2}$	…	$p_i^{\,j}$	…	$p_i^{\,k}$	P_i
…	…	…	…	…	…	…	…
x_n	$p_n^{\,1}$	$p_n^{\,2}$	…	$p_n^{\,j}$	…	$p_n^{\,k}$	P_n
	P^1	P^2	…	P^j	…	P^k	…

(*)

Очевидно $\sum_{i=1}^{n}\sum_{j=1}^{k} p_i^j = 1$

Если просуммировать все p_i^j в i–й строке, то получим

$$\sum_{j=1}^{k} p_i^j = P_i$$

вероятность того, что случайная величина ξ примет значение x_i. Аналогично, если просуммировать все p_i^j в j–м столбце, то получим

$$\sum_{i=1}^{n} p_i^j = P^j$$

вероятность того, что η принимает значение y^j.

Соответствие $x_i \to P_i$ $(i = 1,2,\ldots,n)$ определяет закон распределения ξ, также как соответствие $y^j \to P^j$ $(j = 1,2,\ldots,k)$ определяет закон распределения случайной величины η.

Очевидно $M_\xi = \sum_{i=1}^{n} x_i P_i$, $M_\eta = \sum_{j=1}^{k} y^j P^j$.

Раньше мы говорили, что случайные величины ξ и η независимы, если

$$p_i^j = P_i \cdot P^j \quad (i=1,2,\ldots,n;\ j=1,2,\ldots,k).$$

Если это не выполняется, то ξ и η зависимы.

В чем проявляется зависимость случайных величин ξ и η и как ее выявить из таблицы?

Рассмотрим столбец y^1. Каждому числу x_i поставим в соответствие число

$$p_{i/1} = \frac{p_i^1}{P_1} \qquad (1)$$

которое будем называть условной вероятностью $\xi = x_i$ при $\eta = y^1$. Обратите внимание на то, что это не вероятность P_i события $\xi = x_i$, и сравните формулу (1) с уже известной формулой условной вероятности $P(A/B) = \frac{P(A \cap B)}{P(B)}$.

Соответствие

$$x_i \to p_{i/1},\ (i=1,2,\ldots,n)$$

будем называть условным распределением случайной величины ξ при η=y^1. Очевидно $\sum_{i=1}^{n} p_{i/1} = 1$.

Аналогичные условные законы распределения случайной величины ξ можно построить при всех остальных значениях η, равных y^2; y^3,…, y^n ,ставя в соответствие числу x_i условную вероятность $p_{i/j} = \frac{p_i^j}{P_j}$ ($\sum_{i=1}^{n} p_{i/j} = 1$).

В таблице приведён условный закон распределения случайной величины ξ при η=y^j

ξ	x_1	x_2	…	x_i	…	x_n
$p_{i/j}$	$\frac{p_1^j}{P_j}$	$\frac{p_2^j}{P_j}$	…	$\frac{p_i^j}{P_j}$	…	$\frac{p_n^j}{P_j}$

Можно ввести понятие условного математического ожидания ξ при η = y^j

$$M(\xi / \eta = y_j) = \sum_{i=1}^{n} x_i \frac{p_i^j}{P^j} = \frac{1}{P^j} \sum_{i=1}^{n} x_i p_i^j$$

Заметим, что ξ и η равноценны. Можно ввести условное распределение η при ξ=x_i соответствием

$$y^j \to \frac{p_i^j}{P_i} \quad (j = 1,2,\ldots,k)$$

Также можно ввести понятие условного математического ожидания случайной величины η при ξ=x_i :

$$M(\eta / \xi = x_i) = \sum_{j=1}^{k} y^j \frac{p_i^j}{P_i} = \frac{1}{P_i} \sum_{j=1}^{k} y^j p_i^j$$

Из определения следует, что если ξ и η независимы, то все условные законы распределения одинаковы и совпадают с законом распределения ξ (напоминаем, что закон распределения ξ определяется в таблице (*) первым и последним столбцом). При этом очевидно, совпадают все условные математические ожидания $M(\xi/\eta = y^j)$ при $j = 1,2,\ldots,k$, которые равны Mξ.

Если условные законы распределения ξ при различных значениях η различны, то говорят, что между ξ и η имеет место статистическая зависимость.

Пример I. Пусть закон совместного распределения двух случайных величин ξ и η задан следующей таблицей. Здесь, как говорилось ранее, первый и последний столбцы определяют закон распределения случайной величины ξ, а первая и последняя строки – закон распределения случайной величины η.

ξ \ η	1	2	3	
10	1/36	0	0	1/36
20	2/36	1/36	0	3/36
30	2/36	3/36	2/36	7/36
40	1/36	8/36	16/36	25/36
	6/36	12/36	18/36	

Полигоны условных распределений можно изобразить на трехмерном графике (рис. 1).

Здесь явно просматривается зависимость условного закона распределения ξ от величины η.

Пример II. (Уже встречавшийся).

Пусть даны две независимые случайные величины ξ и η с законами распределения

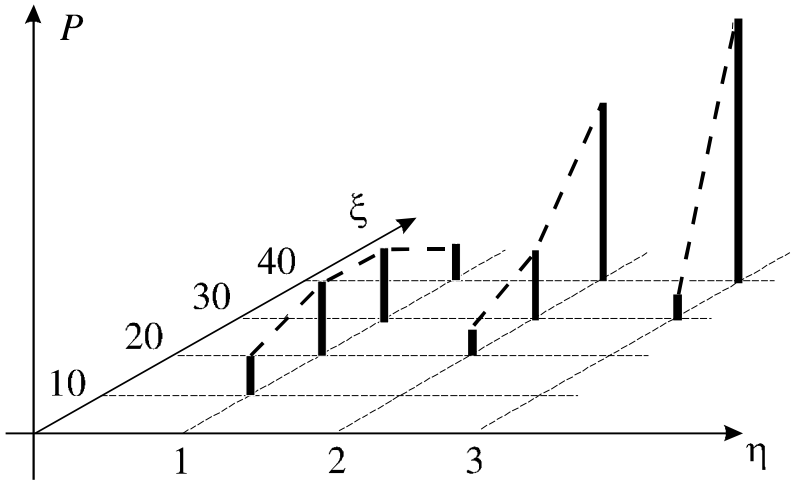

Рис. 1

ξ	0	1
P	1/3	2/3

η	1	2
P	3/4	1/4

Найдем законы распределений случайных величин $\alpha=\xi+\eta$ и $\beta=\xi*\eta$

α	1	2	3
P	3/12	7/12	2/12

β	0	1	2
P	4/12	6/12	2/12

Построим таблицу закона совместного распределения α и β.

β \ α	0	1	2	
1	3/12	0	0	3/12
2	1/12	6/12	0	7/12
3	0	0	2/12	2/12
	4/12	6/12	2/12	

Чтобы получить $\alpha=2$ и $\beta=0$, нужно чтобы ξ приняла значение 0, а η приняла значение 2. Так как ξ и η независимы, то

$$P(\alpha=2;\ \beta=0)= P(\xi=0;\ \eta=2)=P(\xi=0)*P(\eta=2)=1/12.$$

Очевидно также $P(\alpha=3;\ \beta=0)=0$.

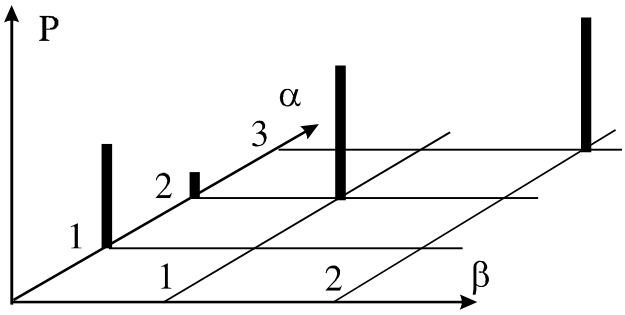

Рис. 2

Построим полигоны условных распределений. Здесь зависимость α от β довольно близка к функциональной: значению β=1 соответствует единственное α=2, значению β=2 соответствует единственное α=3, но при β=0 мы можем говорить лишь, что α с вероятностью $\frac{3}{4}$ принимает значение 1 и с вероятностью $\frac{1}{4}$ – значение 2.

Пример III.

Рассмотрим закон совместного распределения ξ и η, заданный таблицей

ξ \ η	0	1	2	
1	1/30	3/30	2/30	1/5
2	3/30	9/30	6/30	3/5
3	1/30	3/30	2/30	1/5
	1/6	3/6	2/6	

В этом случае выполняется условие $P(\xi=x_i;\ \eta=y^j)=P(\xi=x_i)*P(\eta=y^j)$, i=1,2,3…; j=1,2,3,…

Построим законы условных распределений

ξ	1	2	3
$p_{\eta=1}(\xi) = p_{\eta=2}(\xi) =$ $= p_{\eta=3}(\xi) = p_{\eta=4}(\xi)$	1/5	3/5	1/5

Законы условных распределений не отличаются друг от друга при η=1,2,3 и совпадают с законом распределения случайной величины ξ.

В данном случае ξ и η независимы.

Характеристикой зависимости между случайными величинами ξ и η служит математическое ожидание произведения отклонений ξ и η от их центров распределений (так иногда называют математическое ожидание случайной величины), которое называется коэффициентом ковариации или просто ковариацией.

$$\mathrm{cov}(\xi;\ \eta) = M((\xi - M\xi)(\eta - M\eta))$$

Пусть $\xi = \{x_1, x_2, x_3,\ldots, x_n\}$, $\eta = \{y_1, y_2, y_3,\ldots, y_n\}$. Тогда

$$\mathrm{cov}(\xi;\ \eta)=\sum_{i=1}^{n}\sum_{j=1}^{k}(x_i - M\xi)(y_j - M\eta)P((\xi = x_i)\cap(\eta = y_j)) \qquad (2)$$

Эту формулу можно интерпретировать так. Если при больших значениях ξ более вероятны большие значения η, а при малых значениях ξ более вероятны малые значения η, то в правой части формулы (2) положительные слагаемые доминируют, и ковариация принимает положительные значения.

Если же более вероятны произведения $(x_i - M\xi)(y_j - M\eta)$, состоящие из сомножителей разного знака, то есть исходы случайного эксперимента, приводящие к большим значениям ξ в основном приводят к малым значениям η и наоборот, то ковариация принимает большие по модулю отрицательные значения.

В первом случае принято говорить о прямой связи: с ростом ξ случайная величина η имеет тенденцию к возрастанию.

Во втором случае говорят об обратной связи: с ростом ξ случайная величина η имеет тенденцию к уменьшению или падению.

Если примерно одинаковый вклад в сумму дают и положительные и отрицательные произведения $(x_i - M\xi)(y_j - M\eta)p_i^j$, то можно сказать, что в сумме они будут "гасить" друг друга и ковариация будет близка к нулю. В этом случае не просматривается зависимость одной случайной величины от другой.

Легко показать, что если

$P((\xi = x_i) \cap (\eta = y_j)) = P(\xi = x_i)P(\eta = y_j)$ $(i = 1,2,\ldots,n; j = 1,2,\ldots,k)$,

òî $\operatorname{cov}(\xi; \eta) = 0$.

Действительно из (2) следует

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\left(x_i - M\xi\right)\left(y_j - M\eta\right)P\left(\xi = x_i\right)P\left(\eta = y_j\right) =$$

$$= \sum_{i=1}^{n}\left(x_i - M\xi\right)P\left(\xi = x_i\right) \cdot \sum_{j=1}^{k}\left(y_j - M\eta\right)P\left(\eta = y_j\right) =$$

$$= M\left(\xi - M\xi\right)M\left(\eta - M\eta\right) = 0 \cdot 0 = 0$$

Здесь использовано очень важное свойство математического ожидания: математическое ожидание отклонения случайной величины от ее математического ожидания равно нулю.

Доказательство (для дискретных случайных величин с конечным числом значений).

$$M\left(\xi - M\xi\right) = \sum_{i=1}^{n}\left(x_i - M\xi\right)P\left(x_i\right) = \sum_{i=1}^{n} x_i P\left(x_i\right) - M\xi\sum_{i=1}^{n} P\left(x_i\right) = M\xi - M\xi = 0$$

Ковариацию удобно представлять в виде

$$\operatorname{cov}(\xi; \eta) = M(\xi\eta - \xi M\eta - \eta M\xi + M\xi M\eta) = M(\xi\eta) - M(\xi M\eta) - M(\eta M\xi) + M(M\xi M\eta) =$$
$$= M(\xi\eta) - M\eta M\xi - M\xi M\eta + M\xi M\eta = M(\xi\eta) - M\xi M\eta$$

Ковариация двух случайных величин равна математическому ожиданию их произведения минус произведение математических ожиданий.

Легко доказывается следующее свойство математического ожидания: если ξ и η—независимые случайные величины, то $M(\xi\eta)=M\xi M\eta$. (Доказать самим, используя формулу $M(\xi\eta) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j P(\xi = x_i)P(\eta = y_j)$)

Таким образом, для независимых случайных величин ξ и η $\mathrm{cov}(\xi;\eta)=0$.

Коэффициент корреляции.

Величина cov(ξ;η) зависит от единиц измерения, в которых выражаются ξ и η. (Например, пусть ξ и η—линейные размеры некоторой детали. Если за единицу измерения принять 1 см, то cov(ξ;η) примет одно значение, а если за единицу измерения принять 1 мм, то cov(ξ;η) примет другое, большее значение (при условии cov(ξ;η)≠0)). Поэтому cov(ξ;η) неудобно принимать за показатель связи.

Чтобы иметь дело с безразмерным показателем, рассмотрим случайные величины

$$\xi^* = \frac{\xi - M\xi}{\sqrt{D\xi}} = \frac{\xi - M\xi}{\sigma_\xi};\ \ \eta^* = \frac{\eta - M\eta}{\sqrt{D\eta}} = \frac{\eta - M\eta}{\sigma_\eta}$$

Такие случайные величины называются **нормированными отклонениями** случайных величин ξ и η.

Каждая из случайных величин ξ* и η* имеет центром (математическое ожидание) нуль и дисперсию, равную единице. Приведём доказательство для случайной величины ξ*.

$$M\xi^* = M\left(\frac{\xi - M\xi}{\sigma_\xi}\right) = \frac{1}{\sigma_\xi}\left(M\xi - M\left(M\xi\right)\right) = \frac{1}{\sigma_\xi}\left(M\xi - M\xi\right) = 0$$

$$D\xi^* = D\left(\frac{\eta - M\eta}{\sigma_\xi}\right) = \frac{1}{\sigma_\xi^{\,2}} D\left(\eta - M\eta\right) = \frac{D\eta}{\sigma_\xi^{\,2}} = 1$$

Ковариация ξ* и η* называется **коэффициентом корреляции** случайных величин ξ и η (обозначается $\rho_{\xi\eta}$).

$$\operatorname{cov}(\xi,\eta) = \rho_{\xi\eta} = M\left(\frac{\xi - M\xi}{\sigma_\xi}\cdot\frac{\eta - M\eta}{\sigma_\eta}\right) = \frac{M\left(\left(\xi - M\xi\right)\left(\eta - M\eta\right)\right)}{\sigma_\xi\sigma_\eta} =$$

$$= \frac{\operatorname{cov}(\xi;\eta)}{\sigma_\xi\sigma_\eta} = \frac{M\left(\xi\eta\right) - M\xi M\eta}{\sigma_\xi\sigma_\eta};\ \ \sigma_\xi = \sqrt{D\xi};\ \sigma_\eta = \sqrt{D\eta}.$$

Для независимых ξ и η $\rho_{\xi\eta}$=0, так как в этом случае cov(ξ;η)=0

Обратного заключения сделать нельзя. Случайные величины могут быть связаны даже функциональной зависимостью (каждому значению одной случайной величины соответствует единственное значение другой случайной величины), но коэффициент их корреляции будет равен нулю.

Примеры:

1. Пусть случайная величина ξ симметрично распределена около нуля. Тогда $M\xi$=0. Пусть $\eta=\xi^2$. Тогда $M(\xi\ \eta)=M(\xi^3)$=0, так ξ^3 тоже симметрично распределена около нуля. С другой стороны $M\xi M\eta$=0, так как $M\xi$=0. Таким образом $\rho_{\xi\eta} = \frac{M\left(\xi\eta\right) - M\xi M\eta}{\sigma_\xi\sigma_\eta} = 0$.

2. Пусть закон совместного распределения случайных величин ξ и η задан таблицей

ξ \ η	1	2	
1	1/5	0	1/5
2	0	3/5	3/5
3	1/5	0	1/5
	2/5	3/5	

Проведём вычисления:

$$M\xi = 1\cdot\frac{1}{5} + 2\cdot\frac{3}{5} + 3\cdot\frac{1}{5} = 2;\ M\eta = 1\cdot\frac{2}{5} + 2\cdot\frac{3}{5} = \frac{8}{5};$$

$$M\xi\eta = 1\cdot 1\cdot\frac{1}{5} + 2\cdot 2\cdot\frac{3}{5} + 3\cdot 1\cdot\frac{1}{5} = \frac{16}{5};\ M\xi\eta - M\xi M\eta = 0.$$

Отсюда следует, что $\rho_{\xi\eta}=0$. При этом очевидно, что имеет место функциональная зависимость случайной величины η от случайной величины ξ.

Коэффициент корреляции $\rho_{\xi\eta}$ не меняет своей величины, если вместо случайной величины ξ рассматривать случайную величину $\xi_1=\xi+a$ или $\xi_2=k\xi$ (a и

k—постоянные числа, $k > 0$), так как при перемене начала координат или при изменении масштаба величины ξ нормированное отклонение не меняется. Сказанное в равной мере относится и к η.

Вставка! Полезно запомнить формулу

$D(\xi\pm\eta)=D\xi+D\eta+2\mathrm{cov}(\xi;\eta)$

Отсюда следует свойство дисперсии для независимых ξ и η:

$D(\xi\pm\eta)=D\xi+D\eta$

Свойства коэффициента корреляции.

1. $-1\le\rho_{\xi\eta}\le 1$
2. Если $\rho_{\xi\eta}=1$, то $\eta=k\xi+b$, где k и b—константы, $k>0$.
3. Если $\rho_{\xi\eta}=-1$, то $\eta=k\xi+b$, где $k<0$.
4. Если $\eta=k\xi+b$, ($k\neq 0$) или $\xi=k_1\eta+b_1$, то

$\rho_{\xi\eta}=1$ при $k>0$

$\rho_{\xi\eta}=-1$ при $k<0$.

Коэффициент корреляции $\rho_{\xi\eta}$ достигает своих предельных значений –1 и 1 в том и только в том случае, если совместное распределение ξ и η все концентрируется на некоторой прямой в плоскости ξ; η, то есть между ξ и η имеется такая линейная зависимость.

Если $|\rho_{\xi\eta}|<1$, то такой линейной зависимости нет. Все же по мере приближения $|\rho_{\xi\eta}|$ к единице совместное распределение ξ; η имеет тенденцию концентрироваться вблизи некоторой прямой линии и величину $|\rho_{\xi\eta}|$ можно считать мерой близости к полной линейной зависимости между ξ и η.

Пример. Рассчитаем коэффициент корреляции $\rho_{\xi\eta}$ для случайных величин при заданном законе совместного распределения

ξ \ η	1	2	3	
10	1/36	0	0	1/36
20	2/36	1/36	0	3/36
30	2/36	2/36	2/36	6/36
40	1/36	9/36	16/36	26/36
	6/36	12/36	18/36	

$$M\xi = 10 \cdot \frac{1}{36} + 20 \cdot \frac{3}{36} + 30 \cdot \frac{6}{36} + 40 \cdot \frac{26}{36} \cong 35{,}83$$

$$M\eta = 1 \cdot \frac{6}{36} + 2 \cdot \frac{12}{36} + 3 \cdot \frac{18}{36} \cong 2{,}3$$

$$D\xi = (10 - 35{,}83)^2 \cdot \frac{1}{36} + (20 - 35{,}83)^2 \cdot \frac{3}{36} + (30 - 35{,}83)^2 \cdot \frac{6}{36} +$$

$$+ (40 - 35{,}83)^2 \cdot \frac{26}{36} \cong 57{,}64$$

$$\sigma_\xi \cong 7{,}6$$

$$D\eta = (1 - 2{,}3)^2 \cdot \frac{6}{36} + (2 - 2{,}3)^2 \cdot \frac{12}{36} + (3 - 2{,}3)^2 \cdot \frac{18}{36} \cong 0{,}556$$

$$\sigma_\eta \cong 0{,}746$$

$$M(\xi\eta) = 10 \cdot 1 \cdot \frac{1}{36} + 20 \cdot 1 \cdot \frac{2}{36} + 20 \cdot 2 \cdot \frac{1}{36} + 30 \cdot 1 \cdot \frac{2}{36} + 30 \cdot 2 \cdot \frac{2}{36} +$$

$$+ 30 \cdot 3 \cdot \frac{2}{36} + 40 \cdot 1 \cdot \frac{1}{36} + 40 \cdot 2 \cdot \frac{9}{36} + 40 \cdot 3 \cdot \frac{16}{36} = 86{,}94$$

$$\rho_{\xi\eta} = \frac{86{,}94 - 2{,}3 \cdot 35{,}83}{7{,}6 \cdot 0{,}746} \cong 0{,}8$$

Введем понятие корреляционной зависимости между ξ и η.

Пусть задан закон совместного распределения двух случайных величин ξ и η (как в вышеприведенном примере), и условное математическое ожидание ξ меняется в зависимости от значения η. Тогда принято говорить о корреляционной зависимости ξ от η. Если условное математическое ожидание ξ есть линейная функция от η, то между ξ и η имеется линейная корреляционная связь или зависимость.

Как правило, говоря о корреляционной зависимости, имеют в виду линейную корреляционную зависимость. Если имеется в виду нелинейная корреляционная зависимость, то это особо оговаривают.

Можно дать определение корреляционной зависимости двух случайных величин ξ и η как связи между тенденциями роста ξ и η. Например, между ξ и η существует прямая корреляционная зависимость, если с ростом ξ случайная

величина η имеет тенденцию возрастать. (Это означает, что при больших значениях ξ с большей вероятностью встречаются большие значения η). Если большим значениям ξ ñ большей вероятностью соответствуют меньшие значения η, то есть с ростом ξ случайная величина η имеет тенденцию убывать, говорят, что между ξ и η существует обратная корреляционная зависимость.

Глубина (или теснота) корреляционной зависимости (или связи) характеризуется коэффициентом $\rho_{\xi\eta}$. Чем ближе $|\rho_{\xi\eta}|$ к единице, тем теснее глубина корреляционной зависимости.

Чем ближе зависимость между условным математическим ожиданием ξ и случайной величиной η к линейной, и чем теснее значения ξ группируются около условных математических ожиданий, тем глубже (теснее) корреляционная связь.

Можно говорить о совместном распределении двух непрерывных случайных величин. В большинстве случаев возможен переход от непрерывных случайных величин к совместному распределению двух дискретных случайных величин следующим образом.

Нужно разбить отрезок [*a; b*] изменения случайной величины ξ на равные отрезки $[c_0=a; c_1]$; $[c_1; c_2]$; $[c_2; c_3]$,…,$[c_{n-1}; c_n=b]$. За значение случайной величины ξ принять середину каждого отрезка.

Также надо поступить со случайной величиной η, разбив ее область значений [*e*; *f*] на равные отрезки $[g_0 = e; g_1]$; $[g_1; g_e]$…$[g_{k-1}; g_k=f]$, и приняв за возможные значения η середины отрезков $[g_{k-1}; g_k]$. Таким образом мы получили дискретные случайные величины $\xi^*=\{x_1; x_2; \ldots x_n\}$ и $\eta^*=\{y_1; y_2; \ldots y_k\}$, причем каждой паре $(x_i; y_j)$ ставится в соответствие вероятность

$$P_i^j = P((\xi\in[c_{i-1}; c_i])\cap(\eta\in[g_{i-1}; g_i]))$$

Таким образом мы придем к уже изученному материалу.

Распределение χ^2.

Пусть имеется n независимых случайных величин $\xi_1, \xi_2, ..., \xi_n$, распределенных по нормальному закону с математическим ожиданием, равным нулю, и дисперсией, равной единице. Тогда случайная величина $\chi^2 = \sum_{i=1}^{n} \xi_i^2$ распределена по закону, который называется “распределение χ^2” или “распределение Пирсона”. Очевидно, что она может принимать лишь неотрицательные значения. Число n называется **числом степеней свободы**.

При $n > 1$ график плотности распределения случайной величины χ^2 представляет собой кривую, изображенную на рисунке 1.

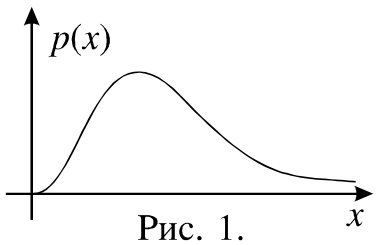

Рис. 1.

Для того, чтобы определить вероятность попадания случайной величины χ^2 в какой-либо промежуток из множества положительных чисел, пользуются таблицей распределения χ^2. Обычно такая таблица позволяет

q / n	0,99	0,975	0,95	...	0,1	0,05	0,01
1	$0,0^3 15$	$0,0^3 98$	$0,0^2 39$	...	2,71	3,84	6,63
...	...	...	...	...	...	...	...
10	2,56	3,25	3,94	...	16,0	18,3	23,2
...	...	...	...	...	...	...	...

Таблица 1.

по вероятности q и по числу степеней свободы n определить так называемый **квантиль** χ_q^2, если q и χ_q^2 связаны соотношением

$$P(\chi^2 > \chi_q^2) = q.$$

Эта формула означает: вероятность того, что случайная величина χ^2 примет значение, большее чем определенное значение χ_q^2, равна q.

Таблица 1 представляет собой фрагмент таблицы распределения χ^2. Из него видно, что случайная величина χ^2 с 10-ю степенями свободы с

вероятностью $q = 0{,}95$ принимает значение, большее 3,94, а та же величина с одной степенью свободы с вероятностью $q = 0{,}975$ превышает 0,00098.

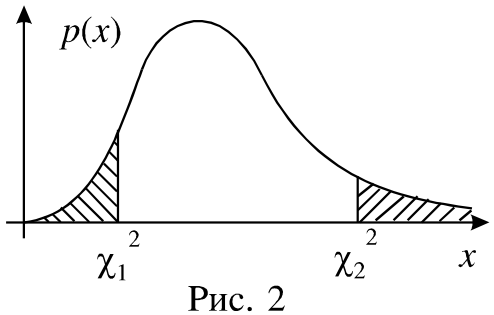

Рис. 2

Задача. Найти интервал (χ_1^2, χ_2^2), в который случайная величина χ^2 с 10-ю степенями свободы попадает с вероятностью, равной 0,9.

Решение. График плотности распределения χ^2 с 10-ю степенями свободы схематично изображен на рисунке 2. Будем считать, что площади заштрихованных областей (правая область не ограничена справа) равны между собой. Примем условия:

$$P(\chi^2 < \chi_1^2) = P(\chi^2 > \chi_2^2) = (1 - 0{,}9)/2 = 0{,}05, \qquad (1)$$

тогда $P(\chi_1^2 < \chi^2 < \chi_2^2) = 0{,}9$.

Равенства (1) сразу позволяют по таблице определить: $\chi_2^2 = 18{,}3$. Для определения левой границы интересующего нас интервала придется воспользоваться очевидным равенством $P(\chi^2 > \chi_1^2) = 0{,}95$. Из таблицы 1. определяем: $\chi_1^2 = 3{,}94$, и теперь можно сформулировать ответ задачи: значение случайной величины χ^2 с вероятностью 0,9 принадлежит интервалу (3,94; 18,3).

Распределение Стьюдента.

Многие задачи статистики приводят к случайной величине вида

$$t = \frac{\xi\sqrt{k}}{\sqrt{\eta}},$$

где ξ и η – независимые случайные величины, причем ξ – нормально распределенная случайная величина с параметрами $M\xi = 0$ и $D\xi = 1$, а η распределена по закону χ^2 с k степенями свободы.

Закон распределения случайной величины t называется **законом распределения Стьюдента** с k степенями свободы.

График плотности распределения для закона Стьюдента схематически изображен на рисунке 3. Кривая плотности распределения схожа с аналогичной кривой для нормального распределения.

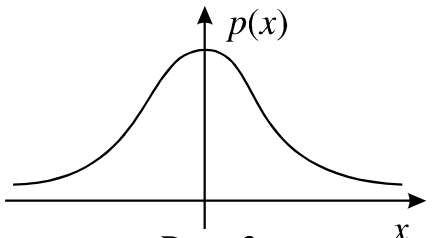

Рис. 3.

Таблицы распределения Стьюдента позволяют при данном числе степеней свободы k по вероятности q определить значение t_q, для которого выполняется соотношение $P(|t| > t_q) = q$. Фрагмент такой таблицы представляет собой таблица 2.

q / k	0,1	0,05	...	0,01	0,005	...
1	6,314	12,71	...	63,57	318	...
...	...	...	...	...	...	...
12	1,782	2,179	...	3,055	3,428	...
...	...	...	...	...	...	...

Таблица 2

Задача. Найти симметричный интервал, в который случайная величина, распределенная по закону Стьюдента с 12-ю степенями свободы, попадает вероятностью 0,9.

Решение. Очевидны соотношения:

$$P(-x < t < x) = P(|t| < x) = 1 - P(|t| \geq x) = 0{,}9.$$

Из последнего равенства следует:

$$P(|t| \geq x) = 0{,}1\ ,\ \ (n = 12).$$

Определяем из таблицы: $x = 1{,}782$. Нестрогое неравенство в скобках в левой части последней формулы нас не должно смущать, так как мы имеем дело с непрерывной случайной величиной, и вероятность того, что она примет конкретное значение, равна нулю.

Задача. Найти значение x из условия $P(t > x) = 0{,}995$, где t – случайная величина, распределенная по закону Стьюдента с 12-ю степенями свободы.

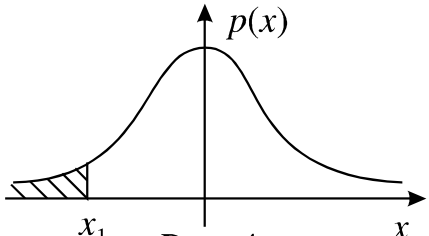

Рис. 4.

Решение. На рисунке 4 изображен график плотности распределения Стьюдента с 12-ю степенями свободы. Вероятность того, что случайная величина примет значение из области справа от точки x_1 равна 0,995 , следовательно в область левее этой точки случайная величина попадает с вероятностью 0,005. Чтобы найти x_1, рассмотрим две симметричные области, изображенные на рисунке 5. Допустим, что в каждой из этих областей значение случайной величины оказывается с вероятностью 0,005. Тогда получаем: $x_1 = -x$, $x_2 = x$, причем x определяется из условия $P(|t| > x) = 0{,}01$. Из таблицы 2 находим: $x = 3{,}055$. Теперь можно выписать ответ задачи:

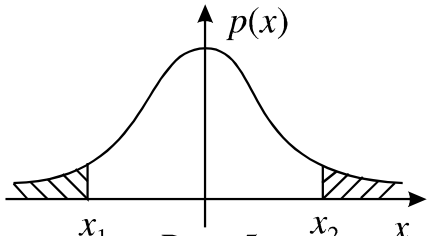

Рис. 5.

$$P(t > -3{,}055) = 0{,}995.$$

Распределение Фишера.

Важные приложения имеет в статистике случайная величина

$$F = \frac{\dfrac{\xi}{k_1}}{\dfrac{\eta}{k_2}} = \frac{k_2\xi}{k_1\eta},$$

где ξ – случайная величина, распределенная по закону χ^2 с k_1 степенями свободы, а η – случайная величина, распределенная по закону χ^2 с k_2 степенями свободы.

Случайная величина F распределена по закону, называемому законом распределения Фишера с k_1 и k_2 степенями свободы. При

заданных числах k_1 и k_2 и по вероятности q по таблице определяется значение F_q такое, что

$$P(F > F_q) = q.$$

Обычно таблицы составляются для значений q, равных 0,05 или 0,01, а иногда для обоих этих значений. Фрагмент такой таблицы представляет собой таблица 3.

k_1 / k_2	1	...	10	...	20	...
1	161,4 647,8	...	241,9 6056	...	248 6209	...
...	...	...	...	...	...	...
10	4,96 10,04	...	2,97 4,85	...	2,77 4,41	...
...	...	...	...	...	...	...

Таблица 3.

В этой таблице в верхней части каждой клетки дается значение F_q при $q = 0{,}05$, а в нижней части — при $q = 0{,}01$.

Математическая статистика.

Основной задачей математической статистики является разработка методов получения научно обоснованных выводов о массовых явлениях и процессах из данных наблюдений и экспериментов. Эти выводы и заключения относятся не к отдельным испытаниям, из повторения которых складывается данное массовое явление, а представляют собой утверждения об общих вероятностных характеристиках данного процесса, то есть о вероятностях, законах распределения, математических ожиданиях, дисперсиях и т. д. Такое использование фактических данных как раз и является отличительной чертой статистического метода.

Пусть мы располагаем сведениями (обычно довольно ограниченными), например, о числе дефектных изделий в изготовленной в определенных условиях продукции или о результатах испытаний материалов на разрушение и т. п. Собранные нами данные могут представлять непосредственный интерес в смысле информации о качестве той или иной партии продукции. Статистические же проблемы возникают тогда, когда мы на основе той же информации начинаем делать выводы относительно более широкого круга явлений. Так например нас может интересовать качество технологического процесса, для чего мы оцениваем вероятность получения в нем дефектного изделия или среднюю долговечность изделия. В этом случае мы рассматриваем собранный материал не ради его самого, а лишь как некую пробную группу или выборку, представляющую только серии из возможных результатов, которые мы могли бы встретить при продолжении наблюдений массового процесса в данной обстановке. Выводы и оценки, основанные на материале наблюдений, отражают случайный состав пробной группы и поэтому считаются приблизительными оценками вероятностного характера. Во многих случаях теория указывает, как наилучшим способом использовать имеющуюся информацию для получения по возможности более точных и надежных характеристик, указывая при этом степень надежности выводов, объясняющуюся ограниченностью запаса сведений.

В математической статистике рассматриваются две основные категории задач: оценивание и статистическая проверка гипотез. Первая задача разделяется на точечное оценивание и интервальное оценивание параметров распределения. Например может возникнуть необходимость по наблюдениям получить точечные оценки параметров $M\xi$ и $D\xi$. Если мы хотим получить некоторый интервал, с той или иной степенью достоверности содержащий истинное значение параметра, то это задача интервального оценивания.

Вторая задача – проверка гипотез – заключается в том, что мы делаем предположение о распределении вероятностей случайной величины (например, о значении одного или нескольких параметров функции распределения) и решаем, согласуются ли в некотором смысле эти значения параметров с полученными результатами наблюдений.

Выборочный метод.

Пусть нам нужно обследовать количественный признак в партии экземпляров некоторого товара. Проверку партии можно проводить двумя способами:

1) провести сплошной контроль всей партии;

2) провести контроль только части партии.

Первый способ не всегда осуществим, например, из–за большого числа экземпляров в партии, из–за дороговизны проведения операции контроля, из–за того, что контроль связан с разрушением экземпляра (проверка электролампы на долговечность ее работы).

При втором способе множество случайным образом отобранных объектов называется **выборочной совокупностью** или **выборкой**. Все множество объектов, из которого производится выборка, называется **генеральной совокупностью**. Число объектов в выборке называется **объемом выборки**. Обычно будем считать, что объем генеральной совокупности бесконечен.

Выборки разделяются на **повторные** (с возвращением) и **бесповторные** (без возвращения).

Обычно осуществляются бесповторные выборки, но благодаря большому (бесконечному) объему генеральной совокупности ведутся расчеты и делаются выводы, справедливые лишь для повторных выборок.

Выборка должна достаточно полно отражать особенности всех объектов генеральной совокупности, иначе говоря, выборка должна быть **репрезентативной** (представительной).

Выборки различаются по способу отбора.

1. Простой случайный отбор.

Все элементы генеральной совокупности нумеруются и из таблицы случайных чисел берут, например, последовательность любых 30-ти идущих подряд чисел. Элементы с выпавшими номерами и входят в выборку.

2. Типический отбор.

Такой отбор производится в том случае, если генеральную совокупность можно представить в виде объединения подмножеств, объекты которых однородны по какому–то признаку, хотя вся совокупность такой однородности не имеет (партия товара состоит из нескольких групп, произведенных на разных предприятиях). Тогда по каждому подмножеству проводят простой случайный отбор, и в выборку объединяются все полученные объекты.

3. Механический отбор.

Отбирают каждый двадцатый (сотый) экземпляр.

4. Серийный отбор.

В выборку подбираются экземпляры, произведенные на каком–то производстве в определенный промежуток времени.

В дальнейшем под генеральной совокупностью мы будем подразумевать не само множество объектов, а множество значений случайной величины, принимающей числовое значение на каждом из объектов. В действительности генеральной совокупности как множества объектов может и не существовать. Например имеет смысл говорить о множестве деталей, которые **можно произвести**, используя данный технологический процесс. Используя какие–то известные нам характеристики данного процесса, мы можем оценивать параметры этого

несуществующего множества деталей. Размер детали – это случайная величина, значение которой определяется воздействием множества факторов, составляющих технологический процесс. Нас, например, может интересовать вероятность, с которой эта случайная величина принимает значение, принадлежащее некоторому интервалу. На этот вопрос можно ответить, зная закон распределения этой случайной величины, а также ее параметры, такие как $M\xi$ и $D\xi$.

Итак, отвлекаясь от понятия генеральной совокупности как множества объектов, обладающих некоторым признаком, будем рассматривать генеральную совокупность как случайную величину ξ, закон распределения и параметры которой определяются с помощью выборочного метода.

Рассмотрим выборку объема n, представляющую данную генеральную совокупность. Первое выборочное значение x_1 будем рассматривать как реализацию, как одно из возможных значений случайной величины ξ_1, имеющей тот же закон распределения с теми же параметрами, что и случайная величина ξ. Второе выборочное значение x_2 – одно из возможных значений случайной величины ξ_2 с тем же законом распределения, что и случайна величина ξ. То же самое можно сказать о значениях x_3, x_4,..., x_n .

Таким образом на выборку будем смотреть как на совокупность независимых случайных величин ξ_1, ξ_2, ..., ξ_n, распределенных так же, как и случайная величина ξ, представляющая генеральную совокупность. **Выборочные значения x_1, x_2, ..., x_n – это значения, которые приняли эти случайные величины в результате 1-го, 2-го, ..., n-го эксперимента**.

Вариационный ряд.

Пусть для объектов генеральной совокупности определен некоторый признак или числовая характеристика, которую можно замерить (размер детали, удельное количество нитратов в дыне, шум работы двигателя). Эта характеристика – случайная величина ξ, принимающая на каждом

объекте определенное числовое значение. Из выборки объема n получаем значения этой случайной величины в виде ряда из n чисел:

$$x_1, x_2,..., x_n. \qquad (*)$$

Эти числа называются значениями признака.

Среди чисел ряда (*) могут быть одинаковые числа. Если значения признака упорядочить, то есть расположить в порядке возрастания или убывания, написав каждое значение лишь один раз, а затем под каждым значением x_i признака написать число m_i, показывающее сколько раз данное значение встречается в ряду (*):

x_1	x_2	x_3	…	x_k
m_1	m_2	m_3	…	m_k

то получится таблица, называемая **дискретным вариационным рядом**. Число m_i называется частотой i-го значения признака.

Очевидно, что x_i в ряду (*) может не совпадать с x_i в вариационном ряду. Очевидна также справедливость равенства

$$\sum_{i=1}^{k} m_i = n\,.$$

Если промежуток между наименьшим и наибольшим значениями признака в выборке разбить на несколько интервалов одинаковой длины, каждому интервалу поставить в соответствие число выборочных значений признака, попавших в этот интервал, то получим **интервальный вариационный ряд**. Если признак может принимать любые значения из некоторого промежутка, то есть является непрерывной случайной величиной, приходится выборку представлять именно таким рядом. Если в вариационном интервальном ряду каждый интервал $[\alpha_i;\ \alpha_{i+1})$ заменить лежащим в его середине числом $(\alpha_i+\alpha_{i+1})/2$, то получим дискретный вариационный ряд. Такая замена вполне естественна, так как, например, при измерении размера детали с точностью до одного миллиметра всем размерам из промежутка [49,5; 50,5), будет соответствовать одно число, равное 50.

Точечные оценки параметров генеральной совокупности.

Во многих случаях мы располагаем информацией о виде закона распределения случайной величины (нормальный, бернуллиевский, равномерный и т. п.), но не знаем параметров этого распределения, таких как $M\xi$, $D\xi$. Для определения этих параметров применяется выборочный метод.

Пусть выборка объема n представлена в виде вариационного ряда. Назовем **выборочной средней** величину

$$\overline{x} = \frac{x_1 m_1 + x_2 m_2 + \ldots + x_k m_k}{n} = x_1 \frac{m_1}{n} + x_2 \frac{m_2}{n} + \ldots + \frac{m_k}{n}$$

Величина $\omega_i = \frac{m_i}{n}$ называется **относительной частотой** значения признака x_i. Если значения признака, полученные из выборки не группировать и не представлять в виде вариационного ряда, то для вычисления выборочной средней нужно пользоваться формулой

$$\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \ .$$

Естественно считать величину $\overline{x}$ выборочной оценкой параметра $M\xi$. Выборочная оценка параметра, представляющая собой число, называется **точечной оценкой**.

Выборочную дисперсию

$$\sigma^2 = \sum_{i=1}^{k} (x_i - \overline{x})^2 \omega_i = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

можно считать точечной оценкой дисперсии $D\xi$ генеральной совокупности.

Приведем еще один пример точечной оценки. Пусть каждый объект генеральной совокупности характеризуется двумя количественными признаками x и y. Например деталь может иметь два размера – длину и ширину. Можно в различных районах измерять концентрацию вредных веществ в воздухе и фиксировать количество легочных заболеваний населения в месяц. Можно через равные промежутки времени

сопоставлять доходность акций данной корпорации с каким-либо индексом, характеризующим среднюю доходность всего рынка акций. В этом случае генеральная совокупность представляет собой двумерную случайную величину ξ, η. Эта случайная величина принимает значения *x*, *y* на множестве объектов генеральной совокупности. Не зная закона совместного распределения случайных величин ξ и η, мы не можем говорить о наличии или глубине корреляционной связи между ними, однако некоторые выводы можно сделать, используя выборочный метод.

Выборку объема *n* в этом случае представим в виде таблицы, где *i*-тый отобранный объект (*i*= 1,2,...*n*) представлен парой чисел x_i, y_i :

x_1	x_2	...	x_n
y_1	y_2	...	y_n

Выборочный коэффициент корреляции рассчитывается по формуле

$$r_{xy} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\sigma_x \sigma_y}$$

Здесь

$$\overline{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i \ , \ \sigma_x = \sqrt{\sigma_x^{\,2}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \ ,$$

$$\sigma_y = \sqrt{\sigma_y^{\,2}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} \ .$$

Выборочный коэффициент корреляции можно рассматривать как точечную оценку коэффициента корреляции $\rho_{\xi\eta}$, характеризующего генеральную совокупность.

Выборочные параметры $\overline{x}, s_x, r_{xy}$ или любые другие зависят от того, какие объекты генеральной совокупности попали в выборку и различаются от выборки к выборке. Поэтому они сами являются случайными величинами.

Пусть выборочный параметр δ рассматривается как выборочная оценка параметра Δ генеральной совокупности и при этом выполняется равенство

$$M\delta = \Delta.$$

Такая выборочная оценка называется **несмещенной.**

Для доказательства несмещённости некоторых точечных оценок будем рассматривать выборку объема n как систему n независимых случайных величин $\xi_1, \xi_2, \ldots \xi_n$, каждая из которых имеет тот же закон распределения с теми же параметрами, что и случайная величина ξ, представляющая генеральную совокупность. При таком подходе становятся очевидными равенства: $Mx_i = M\xi_i = M\xi$; $Dx_i = D\xi_i = D\xi$ для всех $k = 1,2,\ldots n$.

Теперь можно показать, что выборочная средняя $\bar{x}$ есть несмещенная оценка средней генеральной совокупности или , что то же самое, математического ожидания интересующей нас случайной величины ξ:

$$M\bar{x} = M\,\frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n}\left(M\xi_1 + M\xi_2 + \ldots + M\xi_n\right) = \frac{1}{n}nM\xi = M\xi.$$

Выведем формулу для дисперсии выборочной средней:

$$D\bar{x} = D\,\frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{1}{n^2}\left(D\xi_1 + D\xi_2 + \ldots + D\xi_{n1}\right) = \frac{1}{n^2}nD\xi = \frac{D\xi}{n}.$$

Найдем теперь, чему равно математическое ожидание выборочной дисперсии σ^2. Сначала преобразуем σ^2 следующим образом:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - M\xi + M\xi - \bar{x}\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(\left(x_i - M\xi\right)^2 - 2\left(x_i - M\xi\right)\left(\bar{x} - M\xi\right) + \left(\bar{x} - M\xi\right)^2\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n}(x_i - M\xi)^2 - (\bar{x} - M\xi)^2$$

Здесь использовано преобразование:

$$\sum_{i=1}^{n} 2(x_i - M\xi)(\bar{x} - M\xi) = 2(\bar{x} - M\xi)\sum_{i=1}^{n}(x_i - M\xi) =$$

$$= 2(\bar{x} - M\xi)\left(\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} M\xi\right) = 2(\bar{x} - M\xi)(n\bar{x} - nM\xi) = 2n(\bar{x} - M\xi)^2$$

Теперь, используя полученное выше выражение для величины σ^2, найдем ее математическое ожидание.

$$M\sigma^2 = M\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - M\xi)^2 - (\bar{x} - M\xi)^2\right) =$$

$$= \frac{1}{n}\sum_{i=1}^{n} M(x_i - M\xi)^2 - M(\bar{x} - M\xi)^2 = \frac{1}{n} nD\xi - D\bar{x} =$$

$$= D\xi - \frac{D\xi}{n} = \frac{n-1}{n} D\xi .$$

Так как $M\sigma^2 \neq D\xi$, **выборочная дисперсия не является несмещенной оценкой дисперсии генеральной совокупности**.

Чтобы получить несмещенную оценку дисперсии генеральной совокупности, нужно умножить выборочную дисперсию на $\frac{n}{n-1}$. Тогда получится величина $s^2 = \frac{n}{n-1}\sigma^2$, называемая **исправленной выборочной дисперсией**.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

Пусть имеется ряд несмещенных точечных оценок одного и того же параметра генеральной совокупности. Та оценка, которая имеет наименьшую дисперсию называется **эффективной**.

Полученная из выборки объема n точечная оценка δ_n параметра Δ генеральной совокупности называется **состоятельной**, если она сходится по вероятности к Δ. Это означает, что для любых положительных чисел ε и γ найдется такое число $n_{\varepsilon\gamma}$, что для всех чисел n, удовлетворяющих неравенству $n > n_{\varepsilon\gamma}$ выполняется условие $P\left(\left|\delta_n - \Delta\right| < \varepsilon\right) > 1 - \gamma$.

$\bar{x}$ и s^2 являются несмещёнными, состоятельными и эффективными оценками величин $M\xi$ и $D\xi$.

Интервальные оценки.

Точечные оценки параметров генеральной совокупности могут быть приняты в качестве ориентировочных, первоначальных результатов обработки выборочных данных. Их недостаток заключается в том, что неизвестно, с какой точностью оценивается параметр. Если для выборок большого объема точность обычно бывает достаточной (при условии несмещенности, эффективности и состоятельности оценок), то для выборок небольшого объема вопрос точности оценок становится очень важным.

Введем понятие интервальной оценки неизвестного параметра генеральной совокупности (или случайной величины ξ, определенной на множестве объектов этой генеральной совокупности). Обозначим этот параметр через Δ. По сделанной выборке по определенным правилам найдем числа Δ_1 и Δ_2, так чтобы выполнялось условие:

$$P(\Delta_1 < \Delta < \Delta_2) = P(\Delta \in (\Delta_1; \Delta_2)) = \gamma$$

Числа Δ_1 и Δ_2 называются **доверительными границами**, интервал (Δ_1, Δ_2) — **доверительным интервалом** для параметра Δ. Число γ называется **доверительной вероятностью** или **надежностью** сделанной оценки.

Сначала задается надежность. Обычно ее выбирают равной 0.95, 0.99 или 0.999. Тогда вероятность того, что интересующий нас параметр попал в интервал (Δ_1, Δ_2) достаточно высока. Число $(\Delta_1 + \Delta_2) / 2$ – середина доверительного интервала – будет давать значение параметра Δ с **точностью** $(\Delta_2 - \Delta_1) / 2$, которая представляет собой половину длины доверительного интервала.

Границы Δ_1 и Δ_2 определяются из выборочных данных и являются функциями от случайных величин $x_1, x_2, ..., x_n$, а следовательно – сами случайные величины. Отсюда – доверительный интервал (Δ_1, Δ_2) тоже случаен. Он может покрывать параметр Δ или нет. Именно в таком смысле нужно понимать случайное событие, заключающееся в том, что доверительный интервал покрывает число Δ.

Доверительный интервал для математического ожидания нормального распределения при известной дисперсии.

Пусть случайная величина ξ (можно говорить о генеральной совокупности) распределена по нормальному закону, для которого известна дисперсия $D\xi = \sigma^2$ ($\sigma > 0$). Из генеральной совокупности (на множестве объектов которой определена случайная величина) делается выборка объема n. Выборка $x_1, x_2,..., x_n$ рассматривается как совокупность n независимых случайных величин, распределенных так же как ξ (подход, которому дано объяснение выше по тексту).

Ранее также обсуждались и доказаны следующие равенства:

$Mx_1 = Mx_2 = ... = Mx_n = M\xi$;

$Dx_1 = Dx_2 = ... = Dx_n = D\xi$;

$M\bar{x} = M\xi$;

$D\bar{x} = D\xi/n$;

Достаточно просто доказать (мы доказательство опускаем), что случайная величина $\bar{x}$ в данном случае также распределена по нормальному закону.

Обозначим неизвестную величину $M\xi$ через a и подберем по заданной надежности γ число $d > 0$ так, чтобы выполнялось условие:

$$P(|\bar{x} - a| < d) = \gamma \qquad (1)$$

Так как случайная величина $\bar{x}$ распределена по нормальному закону с математическим ожиданием $M\bar{x} = M\xi = a$ и дисперсией $D\bar{x} = D\xi/n = \sigma^2/n$, получаем:

$$P(|\bar{x} - a| < d) = P(a - d < \bar{x} < a + d) =$$

$$= \Phi\left(\frac{a + d - a}{\frac{\sigma}{\sqrt{n}}}\right) - \Phi\left(\frac{a - d - a}{\frac{\sigma}{\sqrt{n}}}\right) = 2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right)$$

Осталось подобрать d таким, чтобы выполнялось равенство $2\Phi\left(\frac{d\sqrt{n}}{\sigma}\right)=\gamma$ или $\Phi\left(\frac{d\sqrt{n}}{\sigma}\right)=\frac{\gamma}{2}$.

Для любого $\gamma \in [0;1]$ можно по таблице найти такое число t, что $\Phi(t)=\gamma/2$. Это число t иногда называют **квантилем**.

Теперь из равенства

$$\frac{d\sqrt{n}}{\sigma}=t$$

определим значение d: $d=\frac{\sigma t}{\sqrt{n}}$.

Окончательный результат получим, представив формулу (1) в виде:

$$P\left(\bar{x}-\frac{\sigma t}{\sqrt{n}}<a<\bar{x}+\frac{\sigma t}{\sqrt{n}}\right)=\gamma\,.$$

Смысл последней формулы состоит в следующем: с надежностью γ доверительный интервал

$$\left(\bar{x}-\frac{\sigma t}{\sqrt{n}};\bar{x}+\frac{\sigma t}{\sqrt{n}}\right)$$

покрывает неизвестный параметр $a = M\xi$ генеральной совокупности. Можно сказать иначе: точечная оценка $\bar{x}$ определяет значение параметра $M\xi$ с точностью $d=\sigma t/\sqrt{n}$ и надежностью γ.

Задача. Пусть имеется генеральная совокупность с некоторой характеристикой, распределенной по нормальному закону с дисперсией, равной 6,25. Произведена выборка объема n = 27 и получено средневыборочное значение характеристики $\bar{x}$ = 12. Найти доверительный интервал, покрывающий неизвестное математическое ожидание исследуемой характеристики генеральной совокупности с надежностью γ=0,99.

Решение. Сначала по таблице для функции Лапласа найдем значение t из равенства $\Phi(t) = \gamma / 2 = 0{,}495$. По полученному значению t = 2,58 определим точность оценки (или половину длины доверительного

интервала) d: $d = 2{,}5\times 2{,}58 / \sqrt{27} \approx 1{,}24$. Отсюда получаем искомый доверительный интервал: (10,76; 13,24).

Доверительный интервал для математического ожидания нормального распределения при неизвестной дисперсии.

Пусть ξ – случайная величина, распределенная по нормальному закону с неизвестным математическим ожиданием $M\xi$, которое обозначим буквой a . Произведем выборку объема n. Определим среднюю выборочную $\bar{x}$ и исправленную выборочную дисперсию s^2 по известным формулам.

Случайная величина

$$t = \frac{(\bar{x} - a)\sqrt{n}}{s}$$

распределена по закону Стьюдента с n – 1 степенями свободы.

Задача заключается в том, чтобы по заданной надежности γ и по числу степеней свободы n – 1 найти такое число t_γ , чтобы выполнялось равенство

$$P\left(\left|\frac{(\bar{x} - a)\sqrt{n}}{s}\right| < t_\gamma\right) = \gamma \qquad (2)$$

или эквивалентное равенство

$$P\left(\bar{x} - t_\gamma \frac{s}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{s}{\sqrt{n}}\right) = \gamma. \qquad (3)$$

Здесь в скобках написано условие того, что значение неизвестного параметра a принадлежит некоторому промежутку, который и является доверительным интервалом. Его границы зависят от надежности γ , а также от параметров выборки $\bar{x}$ и s.

Чтобы определить значение t_γ по величине γ, равенство (2) преобразуем к виду:

$$P\left(\left|\frac{(\bar{x}-a)\sqrt{n}}{s}\right| \geq t_\gamma\right) = 1-\gamma$$

Теперь по таблице для случайной величины t, распределенной по закону Стьюдента, по вероятности $1-\gamma$ и числу степеней свободы $n-1$ находим t_γ. Формула (3) дает ответ поставленной задачи.

Задача. На контрольных испытаниях 20-ти электроламп средняя продолжительность их работы оказалась равной 2000 часов при среднем квадратическом отклонении (рассчитанном как корень квадратный из исправленной выборочной дисперсии), равном 11-ти часам. Известно, что продолжительность работы лампы является нормально распределенной случайной величиной. Определить с надежностью 0,95 доверительный интервал для математического ожидания этой случайной величины.

Решение. Величина $1-\gamma$ в данном случае равна 0,05. По таблице распределения Стьюдента, при числе степеней свободы, равном 19, находим: $t_\gamma = 2{,}093$. Вычислим теперь точность оценки: $2{,}093\times 121/\sqrt{20} = 56{,}6$. Отсюда получаем искомый доверительный интервал:

$$(1943{,}4;\ 2056{,}6).$$

Доверительный интервал для дисперсии нормального распределения.

Пусть случайная величина ξ распределена по нормальному закону, для которого дисперсия $D\xi$ неизвестна. Делается выборка объема n. Из нее определяется исправленная выборочная дисперсия s^2. Случайная величина

$$\chi^2 = \frac{(n-1)s^2}{D\xi}$$

распределена по закону χ^2 с $n-1$ степенями свободы. По заданной надежности γ можно найти сколько угодно границ ${\chi_1}^2$ и ${\chi_2}^2$ интервалов, таких, что

$$P\left({\chi_1}^2 < \chi^2 < {\chi_2}^2\right) = \gamma \qquad (*)$$

Найдем χ_1^2 и χ_2^2 из следующих условий:

$$P(\chi^2 \le \chi_1^2) = (1 - \gamma)/ 2 \qquad (**)$$

$$P(\chi^2 \ge \chi_2^2) = (1 - \gamma)/ 2 \qquad (***)$$

Очевидно, что при выполнении двух последних условий справедливо равенство (*).

В таблицах для случайной величины χ^2 обычно дается решение уравнения $P(\chi^2 \ge \chi_q^2) = q$. Из такой таблицы по заданной величине q и по числу степеней свободы $n - 1$ можно определить значение χ_q^2. Таким образом, сразу находится значение χ_2^2 в формуле (***).

Для определения χ_1^2 преобразуем (**):

$$P(\chi^2 \ge \chi_1^2) = 1 - (1 - \gamma)/ 2 = (1 + \gamma)/ 2$$

Полученное равенство позволяет определить по таблице значение χ_1^2.

Теперь, когда найдены значения χ_1^2 и χ_2^2, представим равенство (*) в виде

$$P\left(\chi_1^2 < \frac{(n-1)s^2}{D\xi} < \chi_2^2 \right) = \gamma .$$

Последнее равенство перепишем в такой форме, чтобы были определены границы доверительного интервала для неизвестной величины $D\xi$:

$$P\left(\frac{(n-1)s^2}{\chi_2^2} < D\xi < \frac{(n-1)s^2}{\chi_1^2} \right) = \gamma .$$

Отсюда легко получить формулу, по которой находится доверительный интервал для стандартного отклонения:

$$P\left(\frac{\sqrt{(n-1)}s}{\sqrt{\chi_2^2}} < \sqrt{D\xi} < \frac{\sqrt{(n-1)}s}{\sqrt{\chi_1^2}} \right) = \gamma \qquad (****)$$

Задача. Будем считать, что шум в кабинах вертолетов одного и того же типа при работающих в определенном режиме двигателях — случайная величина, распределенная по нормальному закону. Было случайным образом выбрано 20 вертолетов, и произведены замеры уровня шума (в децибелах) в каждом из них. Исправленная выборочная дисперсия измерений оказалась равной 22,5. Найти доверительный интервал, накрывающий неизвестное стандартное отклонение величины шума в кабинах вертолетов данного типа с надежностью 98%.

Решение. По числу степеней свободы, равному 19, и по вероятности (1 – 0,98)/2 = 0,01 находим из таблицы распределения χ^2 величину $\chi_2^2 = 36{,}2$. Аналогичным образом при вероятности (1 + 0,98)/2 = 0,99 получаем $\chi_1^2 = 7{,}63$. Используя формулу (****), получаем искомый доверительный интервал: (3,44; 7,49).

Задачи статистической проверки гипотез.

Статистическая проверка гипотез является вторым после статистического оценивания параметров распределения и в то же время важнейшим разделом математической статистики.

Методы математической статистики позволяют проверить предположения о законе распределения некоторой случайной величины (генеральной совокупности), о значениях параметров этого закона (например $M\xi$, $D\xi$), о наличии корреляционной зависимости между случайными величинами, определенными на множестве объектов одной и той же генеральной совокупности.

Пусть по некоторым данным имеются основания выдвинуть предположения о законе распределения или о параметре закона распределения случайной величины (или генеральной совокупности, на множестве объектов которой определена эта случайная величина). Задача заключается в том, чтобы подтвердить или опровергнуть это предположение, используя выборочные (экспериментальные) данные.

Гипотезы о значениях параметров распределения или о сравнительной величине параметров двух распределений называются **параметрическими гипотезами**.

Гипотезы о виде распределения называются **непараметрическими гипотезами**.

Проверить статистическую гипотезу – это значит проверить, согласуются ли данные, полученные из выборки с этой гипотезой. Проверка осуществляется с помощью **статистического критерия**. **Статистический критерий – это случайная величина, закон распределения которой (вместе со значениями параметров) известен в случае, если принятая гипотеза справедлива**. Этот критерий называют еще **критерием согласия** (имеется в виду согласие принятой гипотезы с результатами, полученными из выборки).

Гипотезу, выдвинутую для проверки ее согласия с выборочными данными, называют **нулевой гипотезой** и обозначают H_0. Вместе с

гипотезой H_0 выдвигается **альтернативная** или **конкурирующая** гипотеза, которая обозначается H_1. Например:

1) H_0: $M\xi = 0$ $\quad$ H_1: $M\xi \neq 0$

2) H_0: $M\xi = 0$ $\quad$ H_1: $M\xi > 0$

3) H_0: $M\xi = 0$ $\quad$ H_1: $M\xi = 2$

Пусть случайная величина K – статистический критерий проверки некоторой гипотезы H_0. При справедливости гипотезы H_0 закон распределения случайной величины K характеризуется некоторой известной нам плотностью распределения $p_K(x)$.

Выберем некоторую малую вероятность α, равную 0,05 , 0,01 или еще меньшую. Определим **критическое значение критерия** $K_{кр}$ как решение одного из трех уравнений, в зависимости от вида нулевой и конкурирующей гипотез:

$$P(K > K_{кр}) = \alpha \qquad (1)$$

$$P(K < K_{кр}) = \alpha \qquad (2)$$

$$P((K < K_{кр1}) \cap (K > K_{кр2})) = \alpha \qquad (3)$$

Возможны и другие уравнения, но они встречаются значительно реже, чем приведенные.

Решение уравнения (1) (то же самое для уравнений (2) и (3)) заключается в следующем: по вероятности α, зная функцию $p_K(x)$, заданную как правило таблицей, нужно определить $K_{кр}$.

Что означает условие (1)?

Если гипотеза H_0 справедлива, то вероятность того, что критерий K превзойдет некоторое значение $K_{кр}$ очень мала – 0,05 , 0,01 или еще меньше, в зависимости от нашего выбора. Если $K_в$ – значение критерия K, рассчитанное по выборочным данным, превзошло значение $K_{кр}$, это означает, что выборочные данные не дают основания для принятия нулевой гипотезы H_0 (например, если α=0,01 , то можно сказать, что произошло событие, которое при справедливости гипотезы H_0 встречается в среднем не чаще, чем в одной из ста выборок). В этом случае говорят,

что **гипотеза H_0 не согласуется с выборочными данными и должна быть отвергнута**. Если $K_в$ не превосходит $K_{кр}$, то говорят, что **выборочные данные не противоречат гипотезе H_0**, и нет оснований отвергать эту гипотезу.

Для уравнения (1) область $K > K_{кр}$ называется **критической** областью. **Если значение $K_в$ попадает в критическую область, то гипотеза H_0 отвергается**.

Для уравнения (1) область $K < K_{кр}$ называется **областью принятия гипотезы**. **Если значение $K_в$ попадает в область принятия гипотезы, то гипотеза H_0 принимается**.

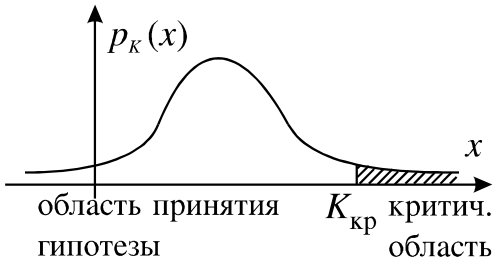

Рис. 1.

Рисунок 1. иллюстрирует решение уравнения (1). Здесь $p_K(x)$ – известная плотность распределения случайной величины K при условии справедливости гипотезы H_0.

Пусть выбрано некоторое малое значение вероятности α, по нему определено значение $K_{кр}$ и по выборочным данным определено значение $K_в$, которое попало в критическую область. В этом случае гипотеза H_0 отвергается, но она может оказаться справедливой, просто случайно произошло событие, которое имеет очень малую вероятность α. В этом смысле α есть вероятность отвержения правильной гипотезы H_0.

Отвержение правильной гипотезы называется **ошибкой первого рода**. Вероятность α называется уровнем значимости. Таким образом **уровень значимости – это вероятность совершения ошибки первого рода**.

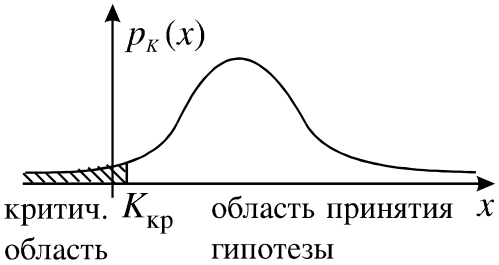

Рис. 2.

Критическая область, полученная для уравнения (1) и приведенная на рисунке 1., называется **правосторонней**.

Уравнение (2) определяет **левосторонюю критическую область**.

Ее изображение приводится на рисунке 2.

Отметим, что каждая из заштрихованных фигур на рисунках 1. и 2. имеет площадь, равную α.

Уравнение (3) определяет **двусторонюю критическую область**. Такая область изображена на рисунке 3. Здесь критическая область состоит из двух частей. В случае двусторонней критической области границы ее частей $K_{кр1}$ и $K_{кр2}$ определяются таким образом, чтобы выполнялось условие:

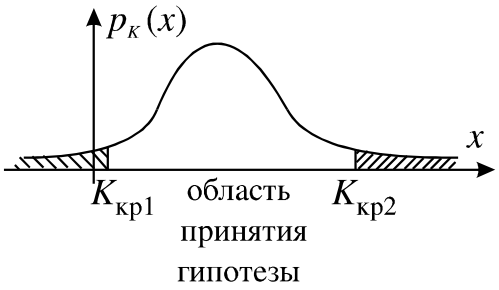

Рис. 3.

$$P(K \le K_{кр}) = P(K \ge K_{кр}) = \alpha / 2.$$

На рисунке 3. площадь каждой из заштрихованных фигур равна $\alpha / 2$.

Вид критической области зависит от того, какая гипотеза выдвинута в качестве конкурирующей.

Чем меньше уровень значимости, тем меньше вероятность отвергнуть проверяемую гипотезу H_0, когда она верна, то есть совершить ошибку первого рода. Но с уменьшением уровня значимости расширяется область принятия гипотезы H_0 и увеличивается вероятность принятия проверяемой гипотезы, когда она неверна, то есть когда предпочтение должно быть отдано конкурирующей гипотезе.

Пусть при справедливости гипотезы H_0 статистический критерий K имеет плотность распределения $p_0(x)$, а при справедливости конкурирующей гипотезы H_1 – плотность распределения $p_1(x)$. Графики этих функций приведены на рисунке 4. При некотором уровне значимости находится критическое значение $K_{кр}$ и правостороняя критическая область. Если значение $K_в$, определенное по выборочным данным, оказывается меньше, чем $K_{кр}$, то гипотеза H_0 принимается. Предположим, что справедлива на самом деле конкурирующая гипотеза H_1. Тогда вероятность попадания критерия в область принятия гипотезы H_0 есть некоторое число β, равное площади фигуры, образованной графиком функции $p_1(x)$ и полубесконечной частью горизонтальной координатной

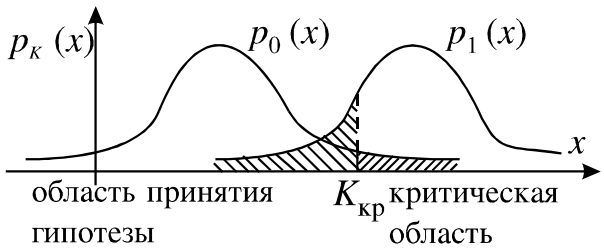

Рис. 4.

оси, лежащей слева от точки $K_{кр}$. Очевидно, что β – это вероятность того, что будет принята неверная гипотеза H_0.

Принятие неверной гипотезы называется ошибкой второго рода. В рассмотренном случае число β – это вероятность ошибки второго рода. **Число** $1 - \beta$, **равное вероятности того, что не совершается ошибка второго рода, называется мощностью критерия**. На рисунке 4 мощность критерия равна площади фигуры, образованной графиком функции $p_1(x)$.и полубесконечной частью горизонтальной координатной оси, лежащей справа от точки $K_{кр}$.

Выбор статистического критерия и вида критической области осуществляется таким образом, чтобы мощность критерия была максимальной.

Проверка статистической гипотезы о математическом ожидании нормального распределения при известной дисперсии.

Пусть имеется нормально распределенная случайная величина ξ,, определенная на множестве объектов некоторой генеральной совокупности. Известно, что $D\xi = \sigma^2$. Математическое ожидание $M\xi$ неизвестно. Допустим, что имеются основания предполагать, что $M\xi = a$, где a – некоторое число (такими основаниями могут быть ограниченные сведения об объектах генеральной совокупности, опыт исследования подобных совокупностей и т. д.). Будем считать также, что имеется другая информация, указывающая на то, что $M\xi = a_1$, где $a_1 > a$.

I. Выдвигаем нулевую гипотезу H_0: $M\xi = a$;

при конкурирующей гипотезе H_1: $M\xi = a_1$.

Делаем выборку объема n: $x_1, x_2,..., x_n$. В основе проверки лежит тот факт, что случайная величина $\bar{x}$ (выборочная средняя) распределена по нормальному закону с дисперсией σ^2/n и математическим ожиданием, равным a в случае справедливости H_0, и равным a_1 в случае справедливости H_1.

Очевидно, что если величина $\bar{x}$ оказывается достаточно малой, то это дает основание предпочесть гипотезу H_0 гипотезе H_1. При достаточно большом значении $\bar{x}$ более вероятна справедливость гипотезы H_1. Задачу можно было бы поставить так: требуется найти некоторое критическое число, которое разбивало бы все возможные значения выборочной средней (в условиях данной задачи это все действительные числа) на два полубесконечных промежутка. При попадании $\bar{x}$ в левый промежуток следовало бы принимать гипотезу H_0, а при попадании $\bar{x}$ в правый промежуток предпочтение следовало бы оказать гипотезе H_1. Однако на самом деле поступают несколько иначе.

В качестве статистического критерия выбирается случайная величина

$$z = \frac{(\bar{x} - a)\sqrt{n}}{\sigma},$$

распределенная по нормальному закону , причем $Mz = 0$ и $Dz = 1$ (это следует из свойств математического ожидания и дисперсии) в случае справедливости гипотезы H_0. Если справедлива гипотеза H_1, то $Mz = a^* = (a_1 - a)\sqrt{n}/\sigma$, $Dz = 1$.

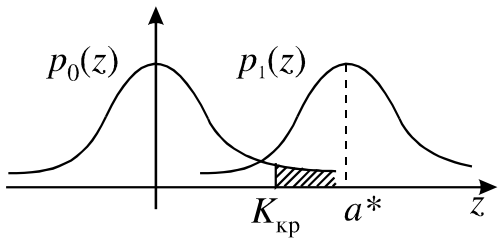

Рис.1.

На рисунке 1. изображены графики $p_0(z)$ и $p_1(z)$ – функций плотности распределения случайной величины z при справедливости гипотез H_0 и H_1, соответственно.

Если величина $\bar{x}$, полученная из выборочных данных, относительно велика, то и величина z велика, что является свидетельством в пользу гипотезы H_1. Относительно малые значения $\bar{x}$ приводят к малым значениям z, что свидетельствует в пользу гипотезы H_0. Отсюда следует, что должна быть выбрана правосторонняя критическая область. По принятому уровню значимости α (например α = 0,05), используя то, что случайная величина z распределена по нормальному закону, определим значение $K_{кр}$ из формулы

$$\alpha = P(K_{кр} < z < \infty) = \Phi(\infty) - \Phi(K_{кр}) = 0{,}5 - \Phi(K_{кр}).$$

Отсюда $\Phi(K_{кр}) = \dfrac{1-2\alpha}{2}$, и осталось воспользоваться таблицей функции Лапласа для нахождения числа $K_{кр}$.

Если величина z, полученная при выборочном значении $\bar{x}$, попадает в область принятия гипотезы $(z < K_{кр})$, то гипотеза H_0 принимается (делается вывод, что выборочные данные не противоречат гипотезе H_0). Если величина z попадает в критическую область, то гипотеза H_0 отвергается.

В данной задаче может быть подсчитана мощность критерия:

$$1 - \beta = \Phi(\infty) - \Phi(K_{\text{ђр}} - \frac{(a_1 - a)}{\sigma}\sqrt{n})$$

Мощность критерия тем больше, чем больше разность $a_1 - a$.

II. Если в предыдущей задаче поставить другое условие:

H_0: $M\xi = a$;

H_1: $M\xi = a_1$, $a_1 < a$,

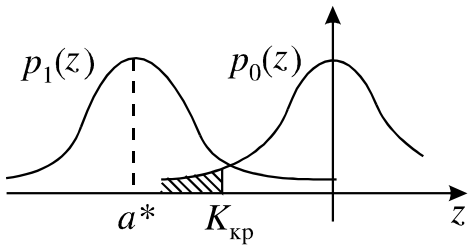

Рис. 2.

то сохранив смысл всех рассуждений, здесь придется рассматривать левостороннюю критическую область, как изображено на рисунке 2. Здесь, как и в предыдущем случае, $a^* = (a_1 - a)\sqrt{n}/\sigma$, а величина $K_{кр}$ определяется из формулы

$$\alpha = P(-\infty < z < K_{кр}) = \Phi(K_{кр}) - \Phi(-\infty) = \Phi(K_{кр}) + \frac{1}{2} .$$

Используя формулу $-\Phi(K_{кр}) = \Phi(-K_{кр})$, получаем:

$$\Phi(-K_{кр}) = \frac{1-2\alpha}{2}.$$

Отметим, что по смыслу задачи здесь $K_{кр}$ – отрицательное число.

Значения z, вычисленные по выборочным данным, превышающие $K_{кр}$, согласуются с гипотезой H_0. Если величина z попадает в критическую область ($z < K_{кр}$), то гипотезу H_0 следует отвергнуть, считая предпочтительной гипотезу H_1.

III. Рассмотрим теперь такую задачу:

H_0: $M\xi = a$;

H_1: $M\xi \neq a$.

В данном случае большие отклонения величины z от нуля в положительную или отрицательную сторону должны приводить к заключению о ложности гипотезы H_0, то есть здесь следует рассматривать двустороннюю критическую область, как изображено на рисунке 3.

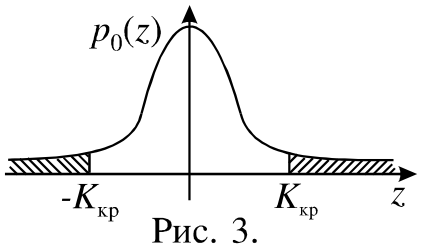

Рис. 3.

Критическое значение $K_{кр}$ определяется с помощью соотношения

$$P(-K_{кр} < z < K_{кр}) = 1 - \alpha = \Phi(K_{кр}) - \Phi(- K_{кр}) = 2\Phi(K_{кр}) .$$

Из этого соотношения следует:

$$\Phi(K_{кр}) = \frac{1-\alpha}{2}.$$

Проверка гипотезы о равенстве дисперсий.

Гипотезы о дисперсии играют очень важную роль в экономико–математическом моделировании, так как величина рассеяния экспериментальных выборочных данных относительно рассчитанных теоретических значений соответствующих параметров, характеризующаяся дисперсией, дает возможность судить о пригодности (адекватности) теории или модели, на основании которой строится теория.

Пусть нормально распределенная случайная величина ξ определена на некотором множестве, образующем генеральную совокупность, а нормально распределенная случайная величина η определена на другом множестве, которое тоже составляет генеральную совокупность. Из обеих совокупностей делаются выборки: из первой – объема n_1, а из второй – объема n_2 (отметим, что объем выборки не всегда можно определить заранее, как например в случае, если он равен количеству рыб, попавших в сеть). По каждой выборке рассчитывается исправленная выборочная дисперсия: s_1^2 для выборки из первой совокупности и s_2^2 для выборки из второй совокупности.

Поставим задачу: с помощью выборочных данных проверить статистическую гипотезу H_0: $D\xi = D\eta$. В качестве конкурирующей гипотезы будем рассматривать идею, заключающуюся в том, что дисперсия той совокупности, для которой исправленная выборочная дисперсия оказалась наибольшей, больше дисперсии другой совокупности. Критерий берется в следующем виде:

$$F = \frac{S^{**}}{S^{*}}.$$

Здесь S^{**}– наибольшая из двух оценок s_1^2 и s_2^2, а S^{*}– наименьшая из тех же двух оценок.

Критерий F распределен по закону Фишера с k_1 и k_2 степенями свободы. Здесь

$$k_1 = n_1 - 1,\ k_2 = n_2 - 1, \text{ если } S^{**} = s_1^2;$$
$$k_1 = n_2 - 1,\ k_2 = n_1 - 1, \text{ если } S^{**} = s_2^2.$$

В этой задаче естественно рассматривать правостороннюю критическую область, так как достаточно большие выборочные значения критерия F свидетельствуют в пользу конкурирующей гипотезы.

При заданном уровне значимости q (обычно q =0,05 или q =0,01) критическое значение $F_{кр}$ определяется из таблицы распределения Фишера. В случае $F > F_{кр}$ гипотеза H_0 отвергается, а в случае $F < F_{кр}$ – принимается.

Пусть два множества некоторых объектов, обладающих количественным признаком, подвергнуты выборочному контролю. Значения количественного признака есть распределенные по нормальному закону случайные величины, которые мы обозначим ξ_1 и ξ_2, соответственно, для первого и для второго множеств. Из первого множества сделана выборка объема n_1=21 и подсчитана исправленная выборочная дисперсия, оказавшаяся равной 0,75. Из второго множества сделана выборка объема n_2=11. Эта выборка дала значение исправленной выборочной дисперсии, равное 0,25. Выдвигаем гипотезу H_0: $D\xi_1 = D\xi_2$. Конкурирующая гипотеза H_1 заключается в том, что $D\xi_1 > D\xi_2$. В данном случае выборочное значение $F_в$ критерия Фишера равно 3. При выбранном уровне значимости $q = 0{,}05$ по числам степеней свободы k_1=20, k_2=10 находим по таблице распределения Фишера $F_{кр}$=2,77. Так как $F_в > F_{кр}$, гипотеза о равенстве дисперсий должна быть отвергнута.

Проверка статистической значимости выборочного коэффициента корреляции.

Проверкой статистической значимости выборочной оценки δ **параметра** Δ **генеральной совокупности называется проверка статистической гипотезы** H_0: $\Delta = 0$, **при конкурирующей гипотезе** H_1: $\Delta \neq 0$. **Если гипотеза** H_0 **отвергается, то оценка** δ **считается статистически значимой**.

Пусть имеются две случайные величины ξ и η, определенные на множестве объектов одной и той же генеральной совокупности, причем обе имеют нормальное распределение. Задача заключается в проверке статистической гипотезы об отсутствии корреляционной зависимости между случайными величинами ξ и η.

H_0: $\rho_{\xi\eta} = 0$;

H_1: $\rho_{\xi\eta} \neq 0$.

Здесь $\rho_{\xi\eta}$ – коэффициент линейной корреляции.

Производится выборка объема n и вычисляется выборочный коэффициент корреляции r. За статистический критерий принимается случайная величина

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

которая распределена по закону Стьюдента с $n - 2$ степенями свободы.

Отметим сначала, что все возможные значения выборочного коэффициента корреляции r лежат в промежутке $[-1;1]$. Очевидно, что относительно большие отклонения в любую сторону значений t от нуля получаются при относительно больших, то есть близких к 1, значениях модуля r. Близкие к 1 значения модуля r противоречат гипотезе H_0, поэтому здесь естественно рассматривать двустороннюю критическую область для критерия t.

По уровню значимости α и по числу степеней свободы $n - 2$ находим из таблицы распределения Стьюдента значение $t_{кр}$. Если модуль выборочного значения критерия $t_в$ превосходит $t_{кр}$, то гипотеза H_0 отвергается и выборочный коэффициент корреляции считается статистически значимым. В противном случае, то есть если $|t_в| < t_{кр}$ и принимается гипотеза H_0, выборочный коэффициент корреляции считается статистически незначимым.